Faculty of Cognitive Sciences and Human Development

Performance Prediction of Compulsory Subjects and Recommendation of Subject Options for China's New College Entrance Examination

Wang Long

Doctor of Philosophy
2026

Performance Prediction of Compulsory Subjects and Recommendation of Subjects Options for China's New College Entrance Examination

Wang Long

A thesis submitted

In fulfillment of the requirements for the degree of Doctor of Philosophy

Faculty of Cognitive Sciences and Human Development
UNIVERSITI MALAYSIA SARAWAK

2026

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

....

Signature 王龙

Name: Wang Long

Matric No.: 22010278

Faculty of Cognitive Sciences and Human Development

Universiti Malaysia Sarawak

Date :25/02/2026

ACKNOWLEDGEMENT

I am deeply grateful to my parents for their unconditional support throughout my doctoral journey. Their love and encouragement have been my greatest source of strength. They stood by me in every moment, ensuring I could focus on my studies without worry. I owe my heartfelt thanks to my wife, Ms. Fang Xiaochun, for her endless understanding and support. Her care for our two young daughters during this time has been nothing short of remarkable. She has taken on the many challenges of parenting alone so I could dedicate myself to my research. Words cannot fully express my appreciation for the sacrifices they have made for me.

I would like to express my sincere thanks to my supervisor, Professor Dr. Chen Chwen Jen, and my co-supervisor, Professor Dr. Teh Chee Siong. Their guidance has been invaluable in shaping this dissertation. They shared their knowledge and provided thoughtful feedback on my work, pushing me to think critically and improve at every stage. I am truly lucky to have had such dedicated supervisors. Their belief in my potential kept me motivated, and their patience and kindness gave me the confidence to persevere through challenges.

Finally, I wish to thank Universiti Malaysia Sarawak (UNIMAS) for the opportunities and resources they provided during my studies. I also extend my gratitude to my workplace, Fuyang No. 3 Middle School in Anhui Province, for their understanding and support. These institutions gave me the foundation and flexibility to pursue my academic goals. At the same time, I would also like to express my gratitude to one of my former students, who is currently studying computer science. This student helped me process a lot of data using the Orange software. Without their help, this journey would not have been possible. I am forever thankful to everyone who contributed to this milestone in my life.

ABSTRACT

This study addresses a critical gap in Educational Data Mining by concurrently predicting performance in China's New College Entrance Examination (NCEE) compulsory subjects and recommending personalized combinations of six optional subjects. Drawing on Bronfenbrenner's ecological framework, we collected data from 1,127 students and 88 teachers at an urban high school across four dimensions: individual, family, school, and social. Continuous predictors were normalized, and categorical variables were transformed into numerical values. The dataset was split 80/20 for training and testing. Four machine learning algorithms: Naïve Bayes (NB), Decision Tree (DT), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs) were evaluated using accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC). Pearson correlations quantified inter subject dependencies. Feature importance analyses revealed that motivation level dominated Chinese performance prediction, followed by teaching method, gender, past Chinese performance, and teacher's self-efficacy. Mathematics predictors centered on test anxiety, parents' education levels, socioeconomic status (SES) and peer relationships, while English hinged on annual family income, parental involvement, and past English performance. NB outperformed all competitors, attaining accuracies of 95.1% for Chinese, 96.4% for Mathematics, and 90.7% for English. Correlation coefficients indicated a weak Chinese-Mathematics association ($r = 0.124$–0.267), a moderate Chinese-English link ($r = 0.308$–0.416), and a moderate Mathematics-English relationship ($r = 0.365$–0.402). From DT outputs, we distilled rules mapping student profiles to optional subject trios. For example, high self-efficacy and strong peer relationships paired with quality Chemistry instruction yielded a "Physics–Chemistry–Biology" recommendation, whereas robust SES and moderate Biology performance suggested "History–Politics–Geography." The above DT rules enable students to optimize their subject options. Limitations include single school sampling and potential regional biases. Future work should replicate across diverse contexts, explore ensemble methods to enhance both accuracy and interpretability, and implement longitudinal follow up.

Keywords: China's New College Entrance Examination (NCEE), compulsory subjects performance prediction, subject options recommendation, Education Data Mining (EDM), important factors, accuracy model, relationship.

Ramalan Prestasi Mata Pelajaran Wajib dan Cadangan Pilihan Mata Pelajaran untuk Peperiksaan Masuk Kolej Baru China

ABSTRAK

Kajian ini menangani jurang kritikal dalam Perlombongan Data Pendidikan dengan meramal prestasi subjek wajib Eksamen Masuk Kolej Baru (NCEE) China dan secara serentak mencadangkan gabungan personal enam subjek pilihan. Berdasarkan kerangka ekologi Bronfenbrenner, kami mengumpul data daripada 1,127 pelajar dan 88 guru di sebuah sekolah menengah bandar merangkumi empat dimensi: individu, keluarga, sekolah, dan sosial. Peramal berterusan dinormalisasi, dan pembolehubah kategori ditukar kepada nilai berangka. Set data dibahagikan 80/20 untuk latihan dan pengujian. Empat algoritma pembelajaran mesin: Naïve Bayes (NB), Decision Tree (DT), Artificial Neural Networks (ANN), dan Support Vector Machines (SVM) dinilai menggunakan ketepatan, precision, recall, skor F1, dan Matthews Correlation Coefficient (MCC). Korelasi Pearson mengukur kebergantungan antara subjek. Analisis kepentingan ciri mendedahkan bahawa tahap motivasi mendominasi peramalan prestasi Bahasa Cina, diikuti oleh kaedah pengajaran, jantina, prestasi Bahasa Cina lepas, dan efikasi kendiri guru. Peramal Matematik berpusat pada keresahan ujian, tahap pendidikan ibu bapa, status sosioekonomi (SES) dan hubungan rakan sebaya, manakala Bahasa Inggeris bergantung pada pendapatan keluarga tahunan, penglibatan ibu bapa, dan prestasi Bahasa Inggeris lepas. NB mengatasi semua pesaing, mencapai ketepatan 95.1% untuk Bahasa Cina, 96.4% untuk Matematik, dan 90.7% untuk Bahasa Inggeris. Pekali korelasi menunjukkan hubungan Bahasa Cina-Matematik yang lemah ($r = 0.124$–0.267), hubungan Bahasa Cina-Bahasa Inggeris sederhana ($r = 0.308$–0.416), dan hubungan Matematik-Bahasa Inggeris sederhana ($r = 0.365$–0.402). Daripada output DT, kami menyaring peraturan yang memprofil pelajar kepada trio subjek pilihan.

Sebagai contoh, efikasi kendiri tinggi dan hubungan rakan sebaya yang kukuh digabungkan dengan pengajaran Kimia yang berkualiti menghasilkan cadangan "Fizik–Kimia–Biologi", manakala SES yang kukuh dan prestasi Biologi sederhana mencadangkan "Sejarah–Politik–Geografi". Peraturan DT di atas membolehkan pelajar mengoptimumkan pilihan subjek mereka. Batasan termasuk pensampelan sekolah tunggal dan bias wilayah yang berpotensi. Kerja masa depan harus mereplikasi dalam konteks yang pelbagai, meneroka kaedah ensemble untuk meningkatkan ketepatan dan kebolehinterpretasian, dan melaksanakan tindak lanjut longitudinal.

Kata kunci: *Peperiksaan Masuk Kolej Baru China (NCEE), ramalan prestasi mata pelajaran wajib, cadangan pilihan mata pelajaran, Perlombongan Data Pendidikan (EDM), faktor penting, model ketepatan, perkaitan.*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AMS	Academic Motivation Scale
ANNs	Artificial Neural Networks
ARFF	Attribute-Relation File Format
ARIMA	Autoregressive Integrated Moving Average
AUC	Area Under Curve
CGPA	Cumulative Grade Point Average
DM	Data Mining
DT	Decision Tree
EDM	Educational Data Mining
EST	Ecological Systems Theory
GPA	Grade Point Average
ICT	Information and Communication Technology
MCC	Matthew's Correlation Coefficient
NB	Naïve Bayes
RF	Random Forest
ROC	Receiver Operating Characteristic
SD	Sustainable Development
SES	Socioeconomic Status
SNA	Social Network Analysis
SNM	Social Network Mining
SVMs	Support Vector Machines
TSES	Teachers' Sense of Efficacy Scale

CHAPTER 1

INTRODUCTION

1.1 Introduction

This study aims to predict China's New College Entrance Examination (NCEE) performance of high school students in compulsory subjects and provide recommendations for students' optional subjects using Educational Data Mining (EDM). This chapter provides an overview of the research background, presents the problem statement, outlines the research objectives, formulates the research questions, establishes the research framework, provides the definitions of terms, highlights the research significance, and delineates the scope of the study.

1.2 Research Background

China's examination tradition started in the Sui Dynasty, which became the Imperial Examination system (Chen et al., 2023). This government's official selection method emphasised merit over position, transforming Chinese education and society. Contemporary "Gaokao", or China's College Entrance Examination (CEE), is a hard, competitive examination crucial for Chinese students. It opens the door to further education and greatly impacts professional prospects. The Gaokao's intensity and relevance reflect China's longstanding esteem for academic performance and its value in personal success (Watkins & Biggs, 1996). Significant changes to CEE have shown how complicated the country's education system is (Gao, 2023; Han, 2022; Hu, 2022). The creation of NCEE, also known as the "new Gaokao", shows a need for a more thorough testing method for getting entry into college (Chen et al., 2020). The NCEE challenges students regarding subject options, and this issue must be addressed urgently (Yi et al., 2022).

1.2.1 Limitation of CEE and the Advantage of NCEE Reform

In China's past, CEE put a lot of weight on scores, which often decided a student's fate. This caused many societal worries about how fair and open it was (Kong, 2020). This method put a lot of weight on the strict division of subjects into arts and sciences and did not pay much attention to what each student desired. Thus, it led to mismatches between students' chosen fields and their real interests, which wasted school resources (Wang, 2021). The CEE system was also argued for not having flexible governance, which made it harder for new ideas and policies to be implemented, especially in an educational environment that was changing quickly (Han & Fu, 2022). This rigidity led to uneven subject options and poor job planning, which disadvantaged students (Chen et al., 2022 November). There were also problems with CEE, like trouble judging a student's qualities as a whole (Liang et al., 2021).

Kong (2020) emphasises the importance of changes introduced in China's new college entrance examination. It called for a fair and thorough testing system that put student growth and equal chances first. The change focuses on job planning, allows students to choose their subjects, and aims to create bold and creative people in line with China's goals (Wang, 2021). The reforms that were put in place offer a variety of optional subjects (Gu, 2023). These help to minimise biases and provide more empowerment to students in making their options of optional subjects. Finally, the change aligns with what the job market needs and emphasises long-term career planning (Chen et al., 2022 November).

1.2.2 The Development of Educational Data Mining (EDM)

EDM is situated at the convergence of pedagogy, statistics, and computer science, as seen in Figure 1.1 (Romero & Ventura, 2013). As a developing field, this discipline utilises data analytic tools to explore extensive collections of educational data to extract significant patterns and insights that might contribute to improving educational outcomes (Okereke, 2019). The main aim of EDM is not only to comprehend these patterns but also to use them in a manner that may lead to concrete enhancements in the field of education (Araka et al., 2022; Calderon-Valenzuela et al., 2022; Mamaril et al., 2022).

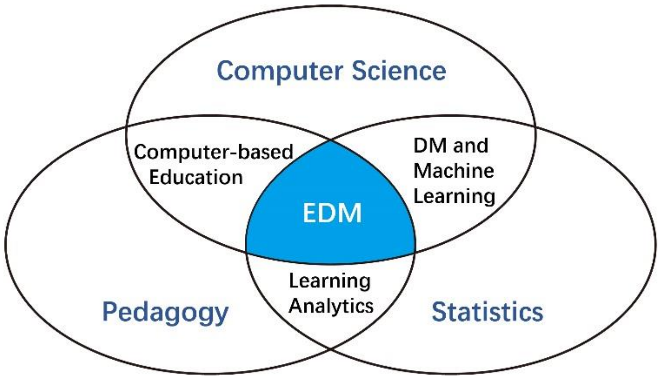

Figure 1.1: Main Areas Related to EDM (Romero & Ventura, 2013)

In the past, two main factors were thought to have caused EDM to grow. First, the digital change made it easier than ever to get educational information (Tavares et al., 2017). This data stems from diverse student engagements with digital tools, encompassing online learning platforms and digital assessment systems (Angrave et al., 2020). This kind of unused data turned out to be a valuable source of information. Second, constant advancements in machine learning techniques provide tools to examine and analyse this data (Marsch & Gustafson, 2013). Machine learning excelled in managing the complexity of educational data due to its capacity to process vast quantities of data and derive trends (Menter et al., 2021; Yağcı, 2022).

EDM can predict students' academic performance, thereby assisting educational policymakers in formulating effective policies (Ben Said et al., 2023). The student performance prediction model demonstrates EDM's significance in assessing possible performance before subject enrolment or examinations, which is essential for personalised education (Zhang et al., 2021). Sarwat et al. (2022) highlight EDM's adaptability in academic activities, particularly in the predictive analysis of student performance data. By making accurate predictions about student performance, stakeholders in educational advancement may devise more effective strategies to enhance educational outcomes (Zhang et al., 2021).

EDM is a way to obtain meaningful information that might impact an organisation and define priority learning requirements for diverse student groups (Algarni, 2016). Based on the analysis of 142 research publications, it is shown that classification methods are the predominant method utilised for evaluating students' performance in EDM (Dol & Jawandhiya, 2023). Decision Tree (DT) is a common kind of classification method algorithms that focus on selecting important features, aiding in global and local explanation (Chatzimparmpas et al., 2023). The C4.5 algorithm, a kind of DT, can accurately classify students to assess their degree of comprehension in a subject with an accuracy rate of 87.50% (Katrina et al., 2019). J48 DT is used to find and name students' learning styles in a virtual learning environment to make a plan for adapting to those styles (Maaliw & Ballera, 2017). DT's rule-based features enable the identification of high-performing students' characteristics (Bilal et al., 2022), allowing for targeted recommendations on their subject options.

1.2.3 Factors to Predict Student Academic Performance

A multitude of variables affect students' academic performance. Predicting student performance is a complex endeavour that necessitates a thorough comprehension of factors most related to and affecting students' performance (Hamoud et al., 2018). These factors encompass individual, family, school, and social dimensions.

The individual dimensions include motivation (Acosta-Gonzaga & Ramirez-Arellano, 2021; Baber, 2020; Borah, 2021; Dunn & Kennedy, 2019; Gumasing & Castro, 2023; Hamid & Singaram, 2016; Khan et al., 2023; Law et al., 2019; Lee et al., 2019; Namoun & Alshanqiti, 2020; Pascoe et al., 2020; Rafiola et al., 2020; Steinmayr et al., 2019; Supervia et al., 2022; Tokan & Imakulata, 2019), self-efficacy (Alemany-Arrebola et al., 2020; Almaiah et al., 2020; Ansari & Khan, 2020;

Fearnley et al., 2022; Hamid & Singaram, 2016; Hayat et al., 2020; Jenal et al., 2022; Liem, 2022; Rafiola et al., 2020; Supervia et al., 2022; Talsma et al., 2021; Wu et al., 2022), past performance (Alghamdi & Rahman, 2023; Alshanqiti, 2020; Alyahyan & Düştegör, 2020; Arashpour et al., 2023; Berberoğlu & Tansel, 2014; Brech & Burnett, 2019; Issah et al., 2023; Mehndiratta & Mehndiratta, 2023; Namoun & Waheed et al., 2020; Rahman,2021; Steinmayr et al., 2019; Supervia et al., 2022), coping strategies (Sharma & Gupta, 2023; Supervia et al., 2022; Liem, 2022), study habits (Islam & Tasnim, 2021; Limniou et al., 2021; Mohamed et al., 2018; Supervia et al., 2022), physical activity (Harveson et al., 2019; Kleszczewska et al., 2018; Sánchez-Hernando et al., 2021; Shantakumar et al., 2022), gender differences (Agasisti et al., 2018; Alemany-Arrebola et al., 2020; Alyahyan & Düştegör, 2020; Cortés Pascual et al., 2019; Hamid & Singaram, 2016; Mohamed et al., 2018; Moldabayev et al., 2013; Nilsen & Gustafsson, 2016), age (Arashpour et al., 2023; Borah, 2021; Cagliero et al., 2021; Cortés Pascual et al., 2019; Fernandes et al., 2019; Mohamed et al., 2018), that directly influence their academic performance. In addition, factors such as emotions (Acosta-Gonzaga & Ramirez-Arellano,2021; Hayat et al., 2020; Luo et al., 2022; Namoun & Alshanqiti, 2020), test anxiety (Akinsola & Nwajei, 2013; Hamid & Singaram, 2016; Jenal et al., 2022; Putwain et al., 2016;), stress (Alemany-Arrebola et al., 2020; Karbownik et al., 2020; Miksza et al., 2021; Pascoe et al., 2020; Supervia et al., 2022), self-esteem (Kleszczewska et al., 2018; Supervia et al., 2022), and cognitive abilities (Cortés Pascual et al., 2019; Galikyan & Admiraal, 2019; Khan et al., 2023; Lee et al., 2019; Peng & Kievit, 2020; Sharma & Gupta, 2023; Wu et al., 2022), are also significant to predict student academic performance.

The family dimensions include parental involvement (Cagliero et al., 2021; Duan et al.,2018; Getie, 2020; Lara & Saracostti, 2019; Leonard et al., 2022; Luo et al., 2022; Mishra, 2020; Mohamed et al., 2018; Nilsen & Gustafsson, 2016; Parmar & Nathans, 2022; Saqib et al., 2018; Tu et al.,2009; Young, 2020), family income (Alghamdi & Rahman, 2023; Leonard et al., 2022; Mamo et al., 2017; Mbogo et al., 2021; Mishra, 2020; Moreira et al., 2019; Rahman, 2021; Rozek et al., 2019; Waheed et al., 2020), and parents’ education levels (Alhadabi & Karpinski, 2020; Echazarra & Radinger, 2019; Leonard et al., 2022; Mehndiratta & Mehndiratta, 2023; Mishra, 2020; Mohamed et al., 2018; Moreira et al., 2019; Rahman, 2021; Suningsih, 2022) play a crucial role in shaping student academic performance.

The school dimensions comprise the quality of education (Gopal et al.,2021; Nilsen & Gustafsson, 2016; Toropova et al., 2019; Wang, 2022), teacher self-efficacy (Toropova et al., 2019; Wang, 2022), teacher qualifications (Assem et al., 2023; Shannag et al., 2013), school climate (Kleszczewska et al., 2018; Ozdemi, 2019; Whittle et al., 2018;), and resources (Almaiah et al., 2020; Mamo et al., 2017). This dimension also includes the impact of teaching methods (Assem et al.,2023; Borah, 2021; Jacobson, 2000; Lekhetho, 2013; Littlejohn, 2020), school leadership (Lekhetho, 2013; Ozdemi, 2019), and learning environment (Abuhassna et al., 2020; Getie, 2020; Law et al., 2019).

The social dimensions encompass the broader social context affecting student academic performance, including socioeconomic status (Alhadabi & Karpinski, 2020; Alyahyan & Düştegör, 2020; Duan et al., 2018; Echazarra & Radinger, 2019; Guo et al., 2023; Issah et al., 2023; Leonard et al., 2022; Mbogo et al., 2021; Mehndiratta & Mehndiratta, 2023; Nilsen & Gustafsson, 2016; Suningsih, 2022; Yeung & Xia, 2023; Young, 2020), peer relationships (Fearnley et al., 2022; Getie, 2020; Lee et al., 2019; Miksza et al., 2021), extracurricular tutoring (Berberoğlu & Tansel, 2014; Chui et al., 2020; Li & He, 2022; Yeung & Xia, 2023; Zhan et al., 2013), social support (Echazarra & Radinger, 2019; Lee et al., 2019; Mishra, 2020), the role of social media (Almaiah et al., 2020; Ansari & Khan, 2020; Dunn & Kennedy, 2019), and extracurricular activities (Agasisti et al., 2018; Mehndiratta & Mehndiratta, 2023).

In summary, this subsection establishes that predicting academic performance requires a multi-dimensional framework. Key determinants are systematically categorized into individual (e.g., motivation, past performance), family (e.g., parental involvement, SES), school (e.g., teaching quality, resources), and social factors (e.g., peer relationships). This taxonomy underscores that academic achievement is not shaped by a single factor but by the complex interplay of a student's entire ecological system, providing a foundational structure for subsequent predictive modeling.

1.2.4 Recommendation of Optional Subjects under NCEE Reform

Under the NCEE system, Chinese, Mathematics and English are compulsory subjects for all students (Zhang et al.,2022). These fundamental subjects serve as the basis of the examination. At present, the importance of Chinese, Mathematics and English has been highlighted, and various measures have been taken to improve the academic performance of these three subjects (Wang, 2021; Zhang, 2020). Furthermore, students in most provinces, except Shanghai and Zhejiang, are given the opportunity to choose from a range of optional subjects such as Physics, Chemistry, Biology, Politics, History, and Geography (Ha & Fu, 2022). These optional subjects enable students to customise their examination according to their aptitudes and passions, offering adaptability within the demanding structure of NCEE. This approach aims to achieve equilibrium between a comprehensive foundation of education and specialised expertise.

The significant shift in NCEE played a pivotal role in allowing students greater autonomy in selecting subjects aligned with their individual interests, skills, and future career aspirations (Yi et al., 2022). To make smart decisions, students had to know everything there was to know about how to get into college, how to use their family’s cultural capital, and how well they could learn. Because of this, there were problems, such as a significant decrease in the number of students opting for more challenging subjects like Physics. In reaction, the “3+3” plan was switched to “3+1+2”, giving students a wider range of subject options. The innovative “3+1+2” educational framework has evolved from the original “3+3” system, now adopted by 23 provinces (Han & Fu, 2022). This model removes the distinction between arts and sciences, requiring students to study three mandatory subjects, namely Chinese, Mathematics and English. Further, they must choose one subject from either Physics or History, and then two from a selection of four - Chemistry, Biology, Politics, and

Geography. This approach represents a significant departure from previous systems, providing a more diverse and customizable educational path. This plan aimed to find a balance between how important subjects were in the national curriculum and how much work students could handle (Xu, 2020). Besides, NCEE reform in Shanghai emphasised increased autonomy in subject options, highlighting the need for clearer recommendations and resources (Li & Tong, 2023).

1.3 Problem Statement

1.3.1 Empirical Gap in the Factors Affecting Academic Performance

Multiple factors affect student academic performance, yet researchers have continually maintained divergent opinions on the importance of certain factors on student academic performance (Atlay et al., 2019; Maghari & Mousa, 2017). Researchers have found that the factor of anxiety has a negative impact on student academic performance. (Jenal et al., 2022; Putwain et al., 2016). However, other researchers have shown that anxiety has an insignificant impact on students' academic performance. (Hamid & Singaram, 2016). Motivation is strongly correlated with positive academic performance (Gumasing & Castro, 2023; Supervia et al., 2022; Tokan & Imakulata, 2019). But motivation did not directly influence learning academic performance in a blended learning setting (Law et al., 2019). Many studies found self-efficacy can improve student academic performance (Almaiah et al., 2020; Fearnley et al., 2022; Jenal et al., 2022; Liem, 2022). On the other hand, self-efficacy alone did not have a significant positive influence on student academic performance (Rafiola et al., 2020). Gender has been identified as a significant factor influencing academic performance (Alyahyan & Düştegör, 2020; Cortés Pascual et al., 2019; Talsma et al., 2021). Interestingly, females did not necessarily translate to a significant difference in academic performance compared to males (Nilsen & Gustafsson, 2016). A systematic investigating major factors to predict students' academic performance hence remains a key research gap. Addressing this knowledge gap, the current study necessitates more investigation in the present study to analyse the factors that predict students' academic performance. This problem of this study needs to identify the most important factors and the highest accuracy model to predict student academic performance in Chinese, Mathematics and English based on NB, DT, ANNs, SVMs algorithms.

Besides, the relationship between Chinese, English and Mathematics is discussed in many publications (Chen & Li, 2008; Cheng et al., 2010; Neville-Barton & Barton, 2005; Peng et al., 2020). A study has shown that the relationship between Chinese and Mathematics is intricate, however there is a substantial interaction between the two (Lu et al., 2022). In general, Chinese-speaking students perform better in mathematics than English-speaking students, and the gap between Chinese and English in mathematics learning is more pronounced, especially for those students who are strong readers but have the poorest mathematical skills (McClung & Arya, 2018). Although complicated, students' mathematical learning is affected by their language of English (Barton & Neville-Barton, 2003). Students who are less proficient in the language of instruction may have greater trouble understanding arithmetic, including exclusion from mathematical conversations and problems utilising visual pictures (Barton & Neville-Barton, 2003). However, a study has pointed out no relationship between English and Mathematics (Mamat, 2016). Proficiency in Chinese may affect English, but the relationship is not simple and direct, and factors like learning background and learners' grade level also play an important role (Yang et al., 2017). Due to the uncertainty of the link between these factors like learning background, learners' grade and academic performance, and the fact that the research did not employ EDM algorithms, data mining may be a useful technique to clarify the inherent relationships (Yale et al., 2017). This is consistent with the needs of this study to identify the relationship between student academic performance in Chinese, Mathematics and English.

1.3.2 Empirical Gap in Existing CEE and NCEE Research

Through investigating the developmental trajectory of CEE and NCEE spanning the years 2014 to 2023 from Google Scholar, it was found that these research around China's Gaokao mainly encompasses a comprehensive investigation of the following dimensions.

a) Historical, Educational, and Psychological Perspectives of the Gaokao System.

Based on the findings of the inquiry, the Gaokao system has established deep historical roots and has had a significant and enduring impact on society over an extended period (Ouyang, 2021; Sha, 2019). Tsegay and Ashraf (2015) highlighted high school educators' significant contribution in preparing for the Gaokao, emphasising their impact on students' academic performance and overall educational journey. The feeling expressed by Liu (2021) resonated with the notion of the cumulative influence exerted by the whole educational system. Yang et al. (2023) emphasised the mental difficulties encountered by students, explicitly drawing attention to the high occurrence of despair and anxiety throughout the Gaokao preparation period, adopting a more personal perspective.

b) Evolving College Admissions Policies and Student Outcomes of Gaokao.

Gu and Yang (2021) provide a comprehensive understanding of the early changes by elucidating their underlying rationales and establishing the framework for subsequent policy modifications. Building upon this inquiry, Huang (2020) and Xu et al. (2023) conducted a comprehensive analysis of the progressive development of these changes over a period, emphasising the subsequent transformations in college admissions and the resulting impact on student outcomes.

c) Comparative Analysis of Gaokao and Global Entrance Examination.

Farley and Yang (2020) and Guo et al. (2021) provide a comprehensive analysis that examines the Gaokao in relation to other international entrance examinations. This comparative study sheds light on the unique features of the Gaokao while also making meaningful connections to similar assessments worldwide. In addition, Yang and Long (2020) conducted a more comprehensive analysis of the differences between the Gaokao and the SAT systems in the United States, providing insights into the advantages and disadvantages associated with each.

d) Socioeconomic and Cultural Dynamics in Gaokao Performance.

Gu (2023) and Chen and Xue (2020) conducted research that examined the differences in Gaokao results, with a specific emphasis on socio-economic aspects. The researchers hypothesised that their socio-economic origins greatly influence students' performance and accomplishments. In addition to considering economic limitations, Howlett (2022) added a cultural lens highlighting the interplay between religious practices, social attitudes, and the pressures associated with Gaokao.

e) Subject Options Autonomy and Influences in Gaokao Outcomes.

Yi et al. (2022) emphasised the significance of student autonomy in selecting subjects, noting that such decisions are notably impacted by factors such as learning efficacy, and the availability of educational resources. Personalised education has expanded the range of available subjects and facilitated a smoother transition between high school and university systems (Xu, 2020). Nevertheless, according to the study conducted by Liang and Zhang (2023) in Zhejiang, it is evident that family background continues to play a crucial role in influencing these decisions and the resulting academic performances. Considering the dynamics, Chen et al. (2022) emphasised the need to implement a complete counselling system that can effectively assist students in making well-informed judgements pertaining to Gaokao.

From 2014 to 2023, researchers investigated China's Gaokao historical importance, social effects, policy changes, foreign comparisons, and factors affecting students' choice of optional subjects. These studies found important issues about the Gaokao history (Sha, 2019; Ouyang, 2021), how teachers prepare students (Tsegay & Ashraf, 2015), the mental challenges students face (Yang et al., 2023), policy changes (Gu & Yang, 2021; Huang, 2020; Xu et al., 2023), international comparisons (Farley & Yang, 2020; Guo et al., 2021; Yang & Long, 2020), and how socio-economic and cultural factors affect performance (Chen & Xue, 2020; Gu, 2023; Howlett, 2022). Some studies also focus on how important it is for students to choose their paths and how family history can affect these options (Chen et al., 2022; Liang & Zhang, 2023; Xu, 2020; Yi et al., 2022;). Studies have briefly mentioned the subject options for Gaokao, but no relevant scholars have given scientific advice on the subject options for it. The critical need arises from the increased decision-making burden on students under the new Gaokao system, who must navigate complex subject combinations without the support of data-driven guidance. Existing studies also do not focus on identifying factors affecting students' Gaokao performance. Hence, the current study looks into factors affecting Gaokao performance and provides recommendations for subject options based on scientific evidence. Then, this problem corresponds to this study to provide recommendation of subject options for each student using DT rules.

1.4 Research Objectives

The primary objective of this study is to identify the most important factors to predict students' academic performance in Chinese, Mathematics, and English using NB, DT, ANNs, SVMs algorithms, as well as to provide some recommendations for students' subject options for NCEE using DT rules.

The specific objectives of this study are outlined below:

i. To identify the most important factors to predict student academic performance in Chinese, Mathematics and English based on NB, DT, ANNs, SVMs algorithms (RO1)

ii. To identify the highest accuracy model in predicting student academic performance in Chinese, Mathematics and English using NB, DT, ANNs, SVMs algorithms (RO2)

iii. To identify the relationship between student academic performance in Chinese, Mathematics and English (RO3)

iv. To provide recommendation of subject options for each student using DT rules (RO4)

1.5 Research Questions

The research questions are created based on the research objectives. RQ1a-RQ1c are designed to achieve the RO1, RQ2a-RQ2c are designed to get the RO2, RQ3 is designed to fulfill the requirements of the RO3, and RQ4a-RQ4f are designed to complete the RO4. The purpose of these questions used to identify best models, understand how predictors vary by subject, generate Decision Tree rules, and recommend subject options in NCEE.

i. What are the most important factors to predict student performance in the Chinese subject based on NB, DT, ANNs, SVMs algorithms (RQ1a)?

ii. What are the most important factors to predict student performance in the Mathematics subject based on NB, DT, ANNs, SVMs algorithms (RQ1b)?

iii. What are the most important factors to predict student performance in the English subject based on NB, DT, ANNs, SVMs algorithms (RQ1c)?

iv. Which predictive model has the highest accuracy in predicting student Chinese performance using NB, DT, ANNs, SVMs algorithms (RQ2a)?

v. Which predictive model has the highest accuracy in predicting student Mathematics performance using NB, DT, ANNs, SVMs algorithms (RQ2b)?

vi. Which predictive model has the highest accuracy in predicting student English performance using NB, DT, ANNs, SVMs algorithms (RQ2c)?

vii. What is the relationship between student performance in Chinese, Mathematics and English (RQ3)?

viii. What are the characteristics of students with high performance in Physics using DT rules (RQ4a)?

ix. What are the characteristics of students with high performance in Chemistry using DT rules (RQ4b)?

x. What are the characteristics of students with high performance in Biology using DT rules (RQ4c)?

xi. What are the characteristics of students with high performance in History using DT rules (RQ4d)?

xii. What are the characteristics of students with high performance in Politics using DT rules (RQ4e)?

xiii. What are the characteristics of students with high performance in Geography using DT rules (RQ4f)?

1.6 Research Framework

The provided schematic, as illustrated in Figure 1.2, presents a comprehensive research framework focused on high school students. This framework encompasses four key dimensions: individual, family, school, and social. These dimensions are analysed to predict student performance in three compulsory subjects: Chinese, Mathematics, and English. The study employs Educational Data Mining (EDM) techniques to identify which of these factors most significantly affect academic performance in these subjects. Furthermore, the study aims to determine the relationship between the performance across these compulsory subjects. A significant aspect of this study is the investigation into the most accurate EDM models for predicting student performance in these key areas. Additionally, the study delves into discovering prominent Decision Tree (DT) rules that characterizes students who exhibit high performance levels in subjects such as Physics, Chemistry, Biology, Politics, History, and Geography. Thus, it can provide recommendation of subject options for each student.

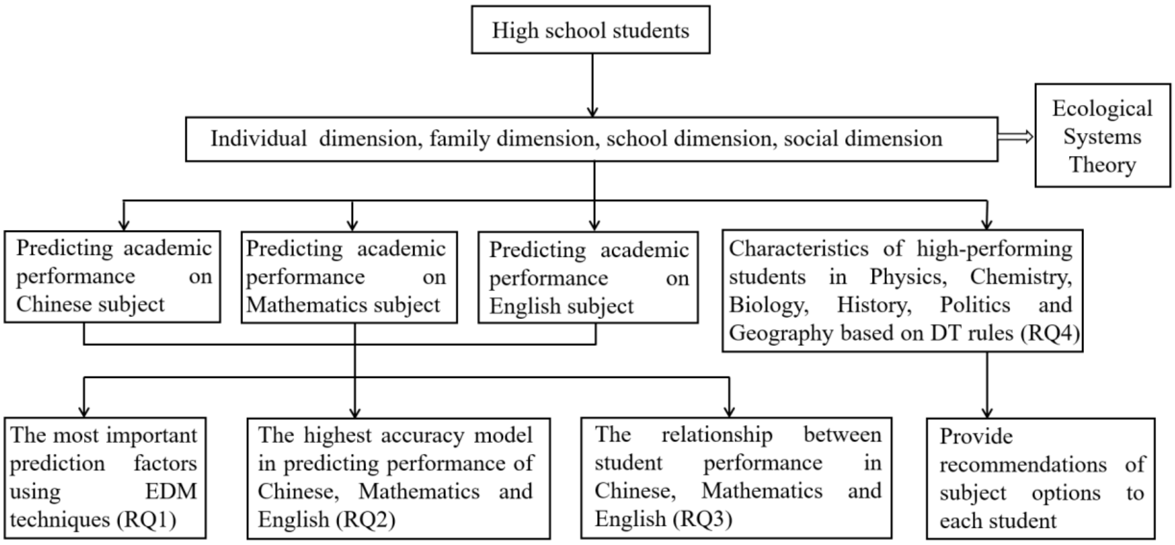

Figure 1.2: Research Framework

1.7 Definitions of Terms

a) China's College Entrance Examination（CEE）

The China's College Entrance Examination (CEE), often referred to as the "old Gaokao", plays a pivotal role as an annual national assessment that determines college eligibility for students (Bai et al., 2013). This critical examination encompasses essential subjects such as Chinese, Mathematics, and English. Additionally, it requires students to make a significant decision between specializing in a science or arts stream, with the stipulation that each student can only select one of these tracks (Zhuoma, 2023). The immense significance of CEE extends beyond academic assessment; it exerts a profound influence on students' academic performance in school and subsequently shapes their future career opportunities.

b) China's New College Entrance Examination（NCEE）

NCEE, a new form of the old CEE or "new Gaokao", was made to meet the changing needs of education. To test students on more than just memory, this change gives them more options in subjects besides the required Chinese, Mathematics, and English subjects (Chen et al., 2021). This increased freedom allows students to customise their academic paths, which could affect their chances of getting into college. Similar with CEE, each student can only choose one combination of sub-subjects in NCEE system.

c) High School Students

Chinese high school students, aged 15-18, undergo a three-year senior secondary education after compulsory junior high, focusing on academic or vocational tracks and preparing for NCEE. High school students will complete three years of intense study before taking NCEE, a highly competitive university entrance examination. This exam significantly influences their future opportunities, leading to immense academic pressure and heavy workloads (Fu, 2024). These pressures are further compounded by limited time and the importance placed on performance. Coping strategies and support systems are essential to mitigate the mental and emotional stress students face (Han, 2024).

d) Student Academic Performance

Student academic performance pertains to the quantifiable academic performance's students attain in their educational endeavours (Supovitz & Comstock, 2023). The study focuses on the scores achieved in NCEE in areas such as Chinese, Mathematics, English, and sub-subjects.

e) Subject Options

Subject options in secondary education allow students to tailor their academic disciplines. Initially, students study a broad curriculum but choose specific subject combinations in the second semester. There are twelve subject options in NCEE. My study focuses on the three subject combinations that the students have chosen: Physics-Chemistry-Biology, Physics-Chemistry-Geography, and History, Politics, and Geography.

f) EDM Techniques

EDM Techniques pertain to the computational methodologies utilised within the field of EDM to analyse and interpret educational data. This study employed algorithms, including Naïve Bayes (NB), Decision Tree (DT), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs) to predict student performance, detection of patterns, and extraction of practical insights from extensive educational datasets. These algorithms are crucial in supporting informed decision-making within the educational domain (Alboaneen et al., 2022; Owusu-Boadu et al., 2021; Ragab et al., 2021). EDM enables the early identification of at-risk students and predicts academic performance.

1.8 Research Significance

The study provides theoretical and practical significance as explained below.

a) Theoretical Significance

Much research on EDM has focused on predicting the performance of higher education and college students (Ding et al., 2019; Ren & Yang, 2020; Yao et al., 2021). EDM prediction model has provided good prediction accuracy of student performance (Wu et al., 2018; Yao et al., 2021), provide insights into appropriate actions to help students completing studies and assist in university decision-making (Cao, 2022; Wang & Wang, 2023). However, as of the end of 2023, no studies have been published in the literature on EDM's prediction of NCEE performance in China. Thus, the study was conducted to fill the gap in EDM on NCEE performance prediction.

Qualitative research examined students' subject options and found that their opinions and satisfaction with subject options difficulties may alter over time (Whiteley et al., 1998). EDM algorithm can reveal the correlation between student performance and subject options, helping them understand their performance and identify possibilities for improvement based on DT (Mkwazu & Yan, 2020), but the study was conducted on university students. The study by Yi et al. (2022) is related to options of optional subjects for NCEE, however it examined the influence of various factors on subject options behaviour, rather than providing recommendations of subject options by examining students' characteristics. Xu (2020) stated that leading K12 extracurricular training companies had used big data analytics to provide NCEE counselling services for students, but it did not explain how big data affects subject options in NCEE. Under the NCEE system, students' final performance depends on their competitors who have chosen the same optional subjects, so they tend to avoid stronger rivals rather than choosing subjects based on their interests and strengths (Chen & Zhu, 2019), which sheds new light on recommendations of subject options. However, as of now there is also no research that uses EDM algorithms, especially DT algorithms, to match high school students with suitable optional subjects by extracting the characteristics of past high performing students of these subjects. Thus, the study fills the gap in NCEE studies on recommendations of subject options.

Previous studies have focused on different factors in predicting student performance. Al-Abyadh et al. (2022) examined students' self-management and perceived self-efficacy, Batool et al. (2023) examined academic and demographic factors, Christensen et al. (2022) examined family, school and community factors, Gore (2016) focused solely on social factors, Levpušček et al. (2013) examined individual and social factors, Li et al. (2019) examined demographics, GPA and parents' education, Lopes et al. (2022) examined individual and school factors, Mishra (2020) focused on social networks, social capital and social support factors, Parr and Bonitz (2015) examined family background and school-related beliefs factors, Pritchard and Wilson (2003) examined emotional and social factors, Singh and Alhulail (2022) examined individual, academic, and socioeconomic factors, Steward focused on school and individual factors, Tadase et al. (2022) examined school quality, individual and family factors, and Waheed et al. (2020) examined individual characteristics. To date, there is only a handful of studies that examine all these dimensions of factors in a single study. Such as a study only considered past academic performance, demographic and psychological factors to predicting students' performance in English and Mathematics using data mining techniques by Roslan and Chen (2023). A systematic review showed that individual student, household context, school community, education systems and macro society factors could predict student performance in the programme for international student assessment (Wang et al., 2023). As this study examines all individual, family, school, and social dimensions simultaneously to predict performance of compulsory subjects using EDM, this study contributes to this pool of literature.

b) Practical Significance

The study produces prediction models that can be used to predict students' NCEE performance in compulsory subjects. This insight can enable teachers and school administrators to carry out appropriate interventions to improve the performance of each student. Another practical implication of this study is the provision of recommendations on subject options. Given the recent revisions implemented in NCEE, students find themselves at a critical juncture when they must make options that have the potential to significantly influence their academic and career trajectories. This study provides objective information to students, allowing them to make informed decisions that match their individual, family, school, and/or social dimensions.

The current study provides accurate recommendations for subject options using DT algorithm, assisting students in making informed judgements that match their abilities and characteristics. It augments the educational experiences of high school students, equipping them with the necessary skills and knowledge to effectively navigate forthcoming hurdles in their academic and professional pursuits.

1.9 Research Scope

The scope is defined by its parameters, which outline the limits of the study concerning its goals, methodology, and potential applications. This study focuses on High School F, which is located in Fuyang City, Anhui Province, China, chosen for its diverse student body with different subject options, which offers a rich dataset to examine the factors affecting student performance. This diversity is key to understanding the nuanced influences on academic outcomes (Cheng, 2020). However, the study's scope is confined to this institution, and its findings should be extrapolated to other settings with caution (Pawlina, 2019), due to the unique characteristics of high school F.

The study is centred on the population of first year at high school F who recently graduated from middle school. The historical datasets from recently graduated students are used to train the prediction model. The model is used to predict the first year students' performance in their upcoming examinations. They are also at a critical point in their academic journey since they have not yet finalised their subject options. The study focuses on first-year high school students to offer timely insights and recommendations that might impact their upcoming subject options, enhancing their academic paths. Besides, first-year high school students have more flexible time to prepare for

Gaokao, and their learning burnout is lower than that of senior two students (Chen et al., 2009). They are more active in choosing subjects and have more opportunities to adjust their subjects (Yuan, 2018).

This study focuses on predictive modelling within the domains of Chinese, Mathematics, and English. The justification for adopting this targeted methodology is two-fold. To begin with, it is important to note that these subjects have a fundamental position within NCEE since they significantly influence a student's total academic performance and, consequently, their qualification for further education (Zou et al., 2016). Furthermore, it is important to note that these subjects are compulsory, regardless of the specific mix of restricted subjects a student choice. The universality of this study guarantees that the findings and suggestions are relevant to all students, irrespective of their chosen subject combinations.

The study utilises the methodology and techniques of EDM to predict student performance (Ragab et al., 2021) and guide subject options. Although EDM provides a methodology based on data analysis and objectivity, it is crucial to acknowledge the inherent constraints associated with this approach. The precision and dependability of the predicts are dependent upon the calibre and comprehensiveness of the accessible data (Yağcı, 2022). Furthermore, whereas EDM can identify patterns and correlations, demonstrating causation necessitates a more comprehensive investigation (Bhegade & Shinde, 2016).

Due to a number of factors, such as the number of teachers, the number of classrooms, and the cancellation of some combinations due to the small number of students choosing them, the present study involves only three combinations of subject options, namely, "Physics, Chemistry and Biology", "Physics, Chemistry and Geography", and "History, Politics and Geography". Hence, only three combinations of subject options are examined. Therefore, further research is needed to explore more possibilities for other subject options.

In summary, the current study utilises EDM approaches to predict student academic performance and provide recommendations for subject options in the context of NCEE, with a specific emphasis on high school students. This study is constrained by its geographical scope, focusing on a specific institution. It places particular emphasis on disciplines like Chinese, Mathematics, and English which have been highlighted (Wang, 2021), since these are compulsory subjects in NCEE and only examines three combinations of subject options. The effectiveness of EDM's data-driven strategy depends upon the data quality used.

1.10 Chapter Summary

This chapter provides an overview of the study context and elucidates the rationale for undertaking the study. Additionally, the document outlines a series of objectives that are then transformed into research questions, serving as a framework for the development of the study. The next chapter provides a comprehensive review of the literature pertaining to this subject.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter aims to examine relevant scholarly literature on the study topic. The subject matter encompasses various aspects such as CEE and NCEE, DM and EDM, methods, algorithms, and tools employed in the field of EDM. Additionally, it explores EDM for academic performance prediction, recommendations of subject options, factors that affect student performance, and the theoretical basis of the study.

2.2 China's College Entrance Examination (CEE) and China's New College Entrance Examination (NCEE)

2.2.1 Introduction to CEE and NCEE

CEE, sometimes called the old Gaokao, has a significant position within China's educational framework (Chen et al., 2020). It functions as the principal means by which individuals who have completed high school in China can enter higher education institutions. The Chinese educational system has implemented substantial reforms to enhance equity and social mobility within CEE, facilitating equitable access to higher education for individuals from diverse socio-economic and geographical contexts (Huang, 2020; Long et al., 2022; Zhang, 2023). Nevertheless, the significant implications associated with this assessment have raised apprehensions over the well-being of students and the accurate evaluation of their capabilities as opposed to just rote learning (Li et al., 2023).

As a mechanism to evaluate candidates' abilities, CEE also plays an important role in helping students grow up and helping the country select talent, maintain education equity, and promote social stability. The annual test significantly influences ten million students' lives, shaping their future paths in school and work (Xu & Lv, 2022). Throughout its history, the field of CEE has seen a series of transformations driven by the changing educational goals and social demands in China (Hu, 2022).

NCEE, often referred to as the new Gaokao, plays a significant function within the Chinese education system as it has the power to shape the academic trajectories of ten million students yearly. In 2014, China launched a new comprehensive reform pilot for CEE. After reforming the CEE subject setting, the total scores of candidates are composed of the scores of three compulsory subjects and three optional subjects (Yi et al., 2022). There are notable distinctions between CEE and NCEE, as shown by thoroughly examining the two in Table 2.1. Throughout history, CEE has been recognised as a demanding and consequential assessment, particularly on memorising facts and demonstrating academic excellence. In recent years, there has been a notable alteration in the configuration and aims of the examination, resulting in the implementation of NCEE. This reform aims to provide a comprehensive evaluation of students, considering their unique preferences, abilities, and objectives. The novel methodology considers not just scholastic performance but also personal aptitudes, inclinations, and ambitions (Li et al., 2018).

Table 2.1: Comparison of CEE and NCEE

Aspects	CEE ("3+3" Model)	NCEE ("3+1+2" Model)
Overview	A standardized national-level examination determining eligibility for higher education institutions.	A reformed version of the CEE, offering students greater flexibility in subject options.
Significance	Pivotal academic milestone for high school students.	Provides a more holistic assessment of students' capabilities.
Model Introduction	"3" stands for the three compulsory subjects - Chinese, Mathematics, and English. Another "3" stands for picking three out of the six optional subjects.	"3" stands for the three compulsory subjects - Chinese, Mathematics, and English; "1" means picking one of Physics and History; and "2" means picking two of the other optional subjects together.
Main Subjects	Chinese, Mathematics, English.	Chinese, Mathematics, English.

Table 2.1 continued

Subject Options	Students choose one of two combinations - Option 1 (science): Physics, Chemistry, and Biology; Option 2 (arts) - History, Politics, and Geography.	4 - Physics, Biology, and Politics; Option 5 - Physics, Biology, and Geography; Option 6 - Physics, Politics, and Geography; Option 7 - History, Politics, and Geography; Option 8: History, Politics, and Chemistry; Option 9 - History, Politics, and Biology; Option 10 - History, Geography, and Chemistry; Option 11 - History, Geography, and Biology; Option 12 - History, Biology, and Chemistry.
Flexibility	Limited to two subject combinations.	Offers a wider range of subject combinations, allowing students to tailor their examination experience.

The reform of NCEE in China has brought about notable transformations in the educational sphere, driven by the country's need to cultivate a proficient workforce that can contribute to its economic advancement (Chen et al., 2022). The curriculum and teaching techniques have been modified to correspond with the goals of the new examination, with a focus on fostering students' personal growth and catering to their interests (Sun, 2023). However, this transition requires significant adaptation from educators, students, and institutions, requiring comprehensive training programs for educators (Zhang et al., 2020; Cai et al., 2019). The significance of parents' educational background, and family monthly income on students' subject options has been emphasised in recent study (Yi et al., 2022). The current study looks into the characteristics of high-performing students for examined optional subjects, which will give insights into the matching between students' characteristics and optional subjects that they potentially perform well.

2.2.2 The Development of CEE and NCEE

Over the ages, China's examination system, notably CEE and NCEE, has evolved. From imperial examinations to its present form, these examination systems have impacted China's education system. These examinations have been constantly updated to satisfy society and education demands (Yu, 2019). This section discusses major reforms in China's examination system, and their effects on students, curriculum, and the educational system.

The examination system in China has a lengthy historical trajectory of more than a millennium, transforming from its inception as imperial examinations to its contemporary manifestation as CEE (Lin & Zhu, 2013). Implementing this approach has played a crucial role in promoting fairness in the selection process of highly skilled personnel and optimising the distribution of resources (Wang, 2018). Jiang et al. (2020) emphasised the many obstacles and adaptations made to the examination system to attain its present configuration.

Since its inception in 1977, CEE has seen notable advancements with the primary objective of augmenting the quality of education. The modifications have been specifically designed to conform to the changing educational policy landscape and tackle the existing regional imbalances effectively. Chen and Feng (2020) underscored the ongoing endeavours to enhance the system, with specific emphasis on conducting periodic evaluations.

The 2014 NCEE reform marked a significant milestone in the development of the examination system. This change was reflective of China's wider national development and reform endeavours (Yang, 2014). The ideals of honesty, democracy, and innovation have played a significant role in shaping student interests, curriculum design, and assessment procedures (Pradana et al., 2020). The primary aim of the 2014 NCEE reform was to promote educational parity (Kong, 2020). The transition was distinguished by a movement from evaluations only based on scores to a comprehensive examination of students (Frame, 2020).

The reform has wide-ranging repercussions. The influence of this phenomenon extends to after-school instruction, the emergence of educational technology, and the widespread availability of online platforms (Sha, 2019). Wang (2021) provided more insights into the impact of the reform on the admissions process, emphasising a shift from a narrow emphasis on standardised examination results to a holistic assessment that considers individuals' preferences and creative aptitudes. The phenomenon has significantly altered the organisational frameworks of secondary educational institutions and the standards used for entrance into professional training colleges.

An outstanding characteristic of the NCEE reform is its prioritisation of student autonomy, individual development, and the promotion of educational informatization. In the study, a comprehensive analysis was provided on the transition from the conventional “3+3” model to the emerging “3+1+2” model (Xu, 2020). It delved into the many issues associated with subject options, resource allocation, and the smooth progression from secondary education to tertiary education (Xu, 2020).

Guo and Yang (2021) conducted a thorough examination of the development trajectory of the NCEE system. The future participation levels were anticipated using the Autoregressive Integrated Moving Average (ARIMA) model. The researchers identified and categorised five major phases in the history of the examination, highlighting its importance as the primary evaluation in China and suggesting possible future developments.

Besides, Liu (2021) studied the evolution of examination in CEE reforms, emphasizing the need for modern approaches in education assessment. Song and Zhang (2019) explored the effect of these reforms on higher education, identifying both benefits and challenges. Bao and Jin (2020) focused on how NCEE changes influenced first-year students, especially in academic performance and major selection. Gong (2020) examined the development of Tianjin’s CEE strategy, showing how policy adapted to local educational needs. Li and Tong (2023) assessed Shanghai’s NCEE rules, noting their positive impact on student autonomy and the persisting issues in subject options and classroom management.

Yang and Long (2020) conducted an in-depth analysis of the CEE system, focusing on its principles of justice, equality, and the associated issues it presents. The researchers conducted a comparative analysis of several international higher education examination methods to evaluate the respective merits and drawbacks. Emphasising the need to adapt their suggestions to the specific circumstances of China, the authors underscored the need to guarantee equity in the admissions process, distribution of information, and availability of opportunities.

An issue of considerable importance within the CEE system is the inequitable allocation of educational resources, particularly between urban and rural regions. Gao (2023) underscored the need to implement a Gaokao policy that promotes equity to provide inclusive education for all individuals, contributing to society's progress. Jing and Liu (2019) analysed the equity of the Gaokao, identifying differences in educational chances across different provinces and between urban and rural areas. It was noted by the authors that, notwithstanding the reform initiatives undertaken by the ministry of education, there are ongoing difficulties in ensuring equitable access to examination chances for all participants.

Chen and Ping et al. (2020) examined the preferential policies implemented in China to enhance educational equity among ethnic minority groups. The study explored these policies' cultural ramifications, obstacles, and potential advantages. The need to enhance these policies was emphasised to provide equitable and unbiased opportunities for ethnic minority populations to pursue higher education.

The growing significance placed on tertiary education in China has resulted in a discernible transformation in students' aspirations for their professional pathways and educational inclinations. Comprehending these changes, particularly within diverse familial origins, is important (Guo et al., 2021).

One notable characteristic of China's Gaokao system is the mandatory selection of a certain academic field by students at the age of 16. The determination of this option is contingent upon several elements, including the economic circumstances of the family and the presence of information asymmetry. Gu (2023) examined the aforementioned factors and argued for the implementation of improved guidance in the process of choosing subjects to maximise the utilisation of educational resources.

Guo et al. (2022) investigated the incorporation of Sustainable Development (SD) principles into the geography curriculum at the high school level in China and its portrayal in NCEE. The study conducted by the researchers yielded valuable findings on the substance, accomplishments, and standardisation of SD in geography education. Specifically, the study shed light on the distinctive attributes of SD-related inquiries found in NCEE.

Long (2023) performed a study that examined the structure and implementation of English elective curriculum in secondary schools in response to the changes introduced by NCEE. The study emphasised the significance of modifying educational goals and instructional approaches to improve students' English language competency and cater to their varied learning requirements.

Lei et al. (2022) investigated the mathematical culture as shown in the mathematics problems in Gaokao spanning 1978 to 2021. The classification system used by the researchers included dividing information into four distinct groups: historical issues, interdisciplinary linkages, social roles, and aesthetics & enjoyment. The research findings indicate an imbalanced concentration, whereby there is a prevailing emphasis on the social ramifications of mathematics, while the examination of its historical development is given less attention. This suggests that the subject matter has changed in correlation with societal transformations.

Liu (2017) highlighted the challenges in offsite CEE policy implementation, revealing geographic disparities in education. Huang (2020) analysed the impact of CEE on county high schools, focusing on holistic child development in rural areas. Han (2022) directed attention to obstacles in educational leadership within the CEE sector in Shandong, proposing a comprehensive approach to students' growth. Yang et al. (2023) examined depression prevalence among students transitioning from Gaokao to college, emphasizing the lack of adequate mental health support.

Wei (2020) conducted a comparative study of educational equality in college entrance examinations between China and the USA, focusing on fairness in examination processes. Sun and Wu (2023) investigated the complexities of timetabling in Chinese high schools post-Gaokao reform, suggesting a two-phase approach for better efficiency. Xu et al. (2023) analysed discrepancies in CEE chemistry exams, noting a limited focus on experimental testing. Howlett (2022) explored the paradoxical nature of religious practices in Gaokao-prep schools, highlighting conflicts between secularism and local beliefs. Lei et al. (2022) examined the mathematical culture in Gaokao questions, focusing on their sociological dimensions and historical context.

The body of research pertaining to NCEE is both complicated and wide. However, it is important to note that there is a dearth of thorough studies primarily focused on subject options aspect of NCEE. Due to the importance of choosing the right optional subjects for students, more studies on this aspect is warranted. NCEE in China has undergone substantial modifications that prioritise comprehensive student assessment and the principles of justice and fairness. The focus of the study is on the element of subject options within the context of NCEE changes. Given the significant ramifications associated with these options, the current study aims to provide specific assistance to students, ensuring that their options are based on existing knowledge and characteristics, and in harmony with their goals.

In summary, navigating the dynamic landscape of NCEE illuminates its profound influence on students' academic pathways. The NCEE's progression (Jiang et al., 2020), from a rigid test to a more inclusive system aligns with contemporary educational reforms to balance regional disparities (Chen & Feng, 2020). The reforms underscore the importance of subject options in shaping students' futures (Gu, 2023), calling for study that can provide insights for students to make informed subject options.

2.3 Data Mining (DM)

DM refers to the methodical examination and interpretation of vast data sets through automated or semi-automated techniques, aimed at uncovering insightful patterns and principles (Tsui et al., 2023). Its versatile applications span multiple domains, including predicting student satisfaction, detecting healthcare anomalies, analysing social media impact, enabling precision marketing, advancing translational neuroscience, improving drought prediction, assessing financial performance, mining social networks, and optimizing medical business operations. The proliferation of DM across several industries has been apparent in recent scholarly investigations. DM has been seen as a highly influential instrument that has the potential to bring about significant changes across all industries and sectors.

a) DM for Predicting Student Satisfaction and Enhancing Online Education. The relevance of Abdelkader et al.'s (2022) study lies in its focus on education during the COVID-19 period, namely in its contribution to predicting student satisfaction and improving online teaching techniques in educational institutions. Through assessing feedback, the researchers identified crucial characteristics that significantly determine the level of success achieved in online learning experiences.

b) DM for Anomaly Detection in Healthcare. Subsequently, Razaque et al. (2022) proposed a novel matrix profile designed for multivariate time series DM within the healthcare industry, explicitly emphasising the identification of anomalies. The study conducted by the authors demonstrated the potential of DM techniques in identifying latent patterns within extensive datasets, leading to enhanced predictive capabilities and the ability to detect abnormalities in time series data.

c) DM for Analysis of Social Media and Its Impact on Audience Attitude. In the context of social media, Hagemann and Abramova (2023) conducted a study using DM techniques to analyse a dataset consisting of over three million tweets related to the 2020 US presidential elections. The results of their study underscored the significant impact that both emotive and cognitive signals have on moulding audience attitude and engagement.

d) DM for Intersection of Business and Precision Marketing. Yang (2023) demonstrated the practicality of DM in precision marketing inside the home broadband industry, as seen in the business sector. Using these methodologies, the researchers developed user-profiles and formulated focused marketing strategies, effectively tackling the obstacles presented by the standardised broadband market. This methodology enhanced the success rates in marketing endeavours and facilitated a more profound comprehension of client preferences.

e) DM for Bioinformatics and Translational Neuroscience. The field of bioinformatics, which involved the use of computational methods to analyse biological data, had significant implications for the field of translational neuroscience, which focused on translating scientific findings into clinical applications. O'Connor et al. (2023) underscored the significant impact of DM in bioinformatics, particularly emphasising its potential for advancing translational neuroscience. With these methodologies on omics datasets, valuable insights pertaining to the pathogenesis of neurodegenerative disorders had been revealed. Consequently, this facilitated the development of more sophisticated treatment approaches and the implementation of personalised medicine.

f) DM for Agriculture and Drought Prediction. In agriculture, Mohammed et al. (2022) emphasised the significance of DM, particularly in drought prediction within the Mediterranean region. By machine learning techniques, researchers enhanced the accuracy of drought predicts, facilitating the adoption of sustainable agricultural methodologies.

g) DM for Analysis of Banking and Financial Performance. DM has also been advantageous for the banking industry. Ledhem (2022) examined the significance of its contribution to the prediction of the financial performance of Islamic banks in Indonesia, a rapidly growing fintech centre.

h) DM for Social Network and Analysis. Nasution (2022) underscored the significance of Social Network Mining (SNM) within the context of the big data age. By extracting dynamic social networks from diverse sources, the researchers successfully established a connection between data sources and Social Network Analysis (SNA), highlighting the significance of data mining in comprehending the development and behaviour of social networks.

i) DM for Medicine Business and Industry. Fu et al. (2023) conducted an in-depth analysis of the utilization of DM techniques within the medical business sector. The authors emphasised the significance of DM in comprehending industry trends and enhancing the assessment models for financial transformation, ultimately leading to improved profitability and efficiency.

2.4 Educational Data Mining (EDM)

The area of EDM has become significant in utilising Data Mining (DM) techniques to evaluate extensive quantities of student data within educational settings (Abdelkader et al., 2022; Okewu et al., 2021). The concept of EDM should not be conflated with DM, as indicated in Table 2.2. The principal objective of EDM is to tackle educational obstacles effectively, optimise learning methodologies, and gain valuable insights into the learning behaviours exhibited by students (Shafiq et al., 2022). The rise in online learning platforms and the growing focus on individualised education has led to the utilisation of EDM techniques and approaches to comprehend and enhance the student learning environment (Zhang et al., 2021).

Table 2.2: Comparison of DM and EDM (Yağcı, 2022)

Aspects	Data Mining (DM)	Educational Data Mining (EDM)
Purpose	It is used to discover new and potentially useful information or meaningful results from big data.	EDM is an effective tool used to identify hidden patterns in educational data, predict academic performance, and improve the learning/teaching environment.
Application	Applied in many sectors, such as banking, healthcare, and marketing.	Applied primarily within the educational sector, aiming to predict student performance and optimise instructional procedures.
Techniques	Generic and generally applicable techniques can be employed.	Techniques can be customised to effectively meet the distinct issues and features inherent in educational data.
Main Goal	The primary objective frequently revolves around enhancing commercial or operational performance.	The primary objective frequently revolves around the enhancement of student performance prediction and the improvement of educational procedures.

Albreiki et al. (2021) used machine learning methods to guess how well students would do in school. This study looked at how EDM can be used to predict which students would drop out and find the ones who were most likely to do so. The above study, analysing data from 2009 to 2021, applied machine learning specifically to address student dropout. In addition, the study showed how important it was for schools to have student assistance programmes to help students do better in school.

In addition, Fatima et al. (2020) thoroughly looked at the different methods and samples used in EDM. Their study aimed to find the best ways to predict how well students would do in school, spot poor behaviour, put students into groups, and make student models. This study showed how vital prediction accuracy was to judge EDM systems.

Moving on to a more specific use, Molins and García's (2022) work was about using EDM methods to find different self-regulated learning patterns in online learning environments. Additionally, they highlighted that agglomerative hierarchical clustering was the best way to find self-regulated learning traits. The study showed that these profiles could help make learning management systems that support self-regulated learning more successfully by considering how learners behaved.

The DT rules of EDM can identify the characteristics of students with different levels of SPM performance in English and Mathematics by evaluating their previous academic performance, parental education levels, and personal traits like diversity and self-criticism, to predict their likelihood of achieving low, moderate, or high performance (Roslan & Chen, 2023).

EDM makes a big difference in education by giving teachers tools and methods that make it easier to understand how students act, guess what will happen, and improve the learning experience (Johnson et al., 2012; Park, 2020). Because more studies in this area, it might be possible to develop more advanced ways to solve the tough problems that schools face today (Stresau & Steiner, 2020).

In summary, EDM offers a powerful approach to predicting student academic performance (Ben Said et al., 2023) and finding out the characteristics of students with high levels of performance in subjects (Roslan & Chen, 2023). Its strength lies in customizing techniques for educational data, optimizing learning methodologies, and supporting individualised education. This study employs EDM to harness these advantages, specifically aiming to comprehend and predict student academic performance, thereby informing educational strategies and providing recommendations for student subject options.

2.5 Methods and Tools for EDM

2.5.1 Methods for EDM

The use of these methods, which are firmly rooted in strong computational and statistical frameworks, has played a crucial role in transforming many domains. These methods provide a data-centric approach to addressing pedagogical obstacles (Depren et al., 2017; Rienties et al., 2020; Yağcı, 2022). This section explores some prominent methods that have been at the forefront of EDM.

a) Classification

Recent study shows that classification methods may alter education fields. Yağcı (2022) used classification algorithms to predict undergraduate students' academic progress using midterm examination data inside machine learning. Data-driven methods in higher education are needed to detect academic underperformers. Ragab et al. (2021) increased student performance prediction using ensemble and EDM classification. This integration enhanced precision. Thus, EDM categorization improves student performances, reveals new insights, and optimises teaching.

Classification methods have emerged as a cornerstone in current study on precise student performance prediction. The ability to extract complex data into actionable categories aligns with the objective of the current study to predict student performance. The diversity of applications, from Križanić's (2020) DT algorithms in e-learning engagement to Yağcı's (2022) use of midterm data for performance prediction, underscores the adaptability of these methods to educational data. Ragab et al. (2021) further validate the efficacy of classification through EDM to boost prediction accuracy.

b) Regression

Regression analysis plays a crucial role in the field of education by providing valuable insights into predicting student performance and comprehending the diverse array of variables that impact learning performance. Regression analysis was used to anticipate the Grade Point Averages (GPA) of students upon their graduation (Tekin, 2014)). The research emphasised the significant influence of EDM in the field of educational predicting. In agreement with this viewpoint, Yağcı (2022) used machine learning methodologies, such as regression analysis, to predict the educational advancement of undergraduate students. The results of their study emphasise the importance of using data-driven approaches in the field of higher education. Moreover, Rao et al. (2017) emphasised the efficacy of regression in EDM for predicting student placements, hence highlighting the diverse range of applications for this methodology.

Regression analysis, with its robust predictive capabilities, is vital to studies in accurately predicting student academic performance. This analytical technique, by considering a multitude of influencing factors, aligns perfectly with the multifaceted nature of educational data. Echoing the work of Tekin (2014) and Yağcı (2022), who successfully applied regression to predict GPA and educational progression, this method as pivotal in deciphering the complexities of student performance. Furthermore, Rao et al.'s (2017) application in placement prediction solidifies regression's adaptability to various educational contexts. Regression method in EDM face significant limitations. Firstly, they assume a linear relationship between variables, which may not capture complex educational phenomena (Namoun & Alshanqiti, 2020)). Secondly, they are sensitive to outliers and multicollinearity, which can skew results and reduce prediction accuracy (Yağcı, 2022). Finally, regression models struggle with imbalanced datasets, common in education, where some outcomes (like dropout rates) are rare, leading to biased predictions (Wongvorachan et al., 2023). These factors make regression less suitable compared to more robust machine learning algorithms.

c) Association Rules

Association rules play a crucial role as they enable the identification and analysis of concealed patterns and correlations present in educational datasets (Arcinas et al., 2021). The use of this approach yields substantial advantages in comprehending student conduct and predicting academic performance (Batool et al., 2023).

Drawing upon the groundwork, Araka et al. (2022) conducted a study using EDM techniques, namely association rules, to discern distinct patterns of self-regulated learning within digital learning contexts. The study underscores the need to use optimum algorithms, such as agglomerative hierarchical clustering, to identify specific patterns of self-regulated learning shown by students. To explore further applications, Chango et al. (2022) undertook a thorough examination of the current body of work on the incorporation of data in multimodal learning analytics and EDM. The study conducted by the authors highlighted the importance of association rules in the amalgamation of diverse educational data sources, including audio, video, and user logs, to get a comprehensive comprehension of the learning process. Ragab et al. (2021) highlighted the potential use of ensemble approaches and EDM methodologies, specifically association rules, to augment the accuracy of student performance prediction. The study demonstrated the effectiveness of using a variety of classifiers to enhance the accuracy of predicting student outcomes. Association rules in EDM have a vital function as they provide valuable insights that may efficiently transform unprocessed educational data into actionable information, eventually resulting in enhanced learning outcomes (Xia, 2020). Although association rules are useful in uncovering hidden patterns in datasets, it is important to note that they may oversimplify complex educational practises and may not adequately capture causal relationships (Rahman et al., 2022).

In summary, EDM approaches are more than computational tools, they underpin modern educational research (Asad et al., 2023). Engels et al. (2022) claimed that data and decision-making processes improve educational interventions, resulting in informed, targeted, and effective results. Three popular EDM methods are classification, regression, and association rules. Classification is commonly utilised in pattern analysis, student success prediction, and engagement understanding. This claim is corroborated by several studies (Križanić, 2020; Yağcı, 2022). Regression is essential for estimating continuous outcomes. This method is effective in predicting student performance (Tekin, 2014; Yağcı, 2022). Association rules reveal hidden patterns in educational datasets, helping researchers understand student behaviour and predict academic performance (Araka et al., 2022). The current study will use classification method and association rules to get comprehensive insights due to their relevance and efficacy.

2.5.2 Tools for EDM

The field of EDM has had significant advancements due to the emergence of advanced technologies that enable the extraction, analysis, and interpretation of educational data.

a) RapidMiner

RapidMiner is widely recognised as a notable software product in EDM, providing a comprehensive suite of tools specifically designed for various tasks including data preparation, modelling, assessment, and deployment. The usefulness of RapidMiner was shown by Berhanu and Abera (2015) in longitudinal study conducted at Dilla University over five years, from 2009 to 2014. The predictive potential of the platform was shown via the examination of students' academic records. Similarly, a study conducted by Shafiq et al. (2022) explored the difficulties associated with student retention, emphasising the crucial significance of EDM and predictive analytics.

Intakaew and Wattanakasiwich (2022) used the capabilities of RapidMiner to augment the competency of Grade-10 students in interpreting kinematics graphs. The results of their study not only emphasised significant enhancements after the instructional intervention and revealed complex associations between students' attitudes towards the process of learning physics. Huerta et al. (2023) used the RapidMiner platform and utilised the knowledge discovery in databases approach within the context of digital marketing for education. The technique used in their study revealed previously undiscovered patterns within extensive datasets, enhancing decision-making processes and strategic planning.

In their study, Inusah et al. (2022) used RapidMiner Studio 9.10 as a tool for data mining in the context of public schools in Ghana. Using classification and clustering methodologies, the researchers identified specific educational obstacles and put forward recommended approaches for the efficient allocation of resources. In their comprehensive study, Karalić et al. (2023) thoroughly examined open-source data mining tools, particularly emphasising the evaluation of RapidMiner. The frictionless installation, Java-based architecture, and possibility for additional functionality were emphasised.

RapidMiner has shown its proficiency in the field of sentiment analysis. Remali et al. (2022) conducted an analysis of attitudes towards online learning in the context of the COVID-19 pandemic. They used the RapidMiner platform with the Support Vector Machines (SVMs) classifier to achieve a notable accuracy rate of roughly 90.41%. Similarly, the study conducted by Adi et al. (2022) used a combination of RapidMiner tools and the Naïve Bayes approach to assess the feelings around online schooling in Indonesia during the epidemic. Abrori et al. (2022) conducted significant study whereby they used the C4.5 algorithm in conjunction with adaptive boosting on the RapidMiner platform to predict student eligibility for college finance. Impressively, this approach yielded a notable accuracy rate of 98.55%. Gumba et al. (2022) highlighted the use of RapidMiner in the analysis of student admissions to ITE courses. They commended its effectiveness in handling a wide range of data qualities, enabling the generation of well-informed conclusions.

RapidMiner has established a notable presence within the field of EDM. However, using RapidMiner is powerful but cannot handle big datasets, needs parameter settings, may output many rules, and requires user-defined support and confidence levels for reliable results (Mayaza et al., 2023). Besides, it is a commercial software, so that the current study does not use it.

b) Weka

Weka, a widely acknowledged tool in the domain of EDM, is renowned for its extensive collection of machine learning algorithms and data preprocessing utilities (Salihoun, 2020).

In their study, Hussain et al. (2018) used the Weka software to analyse student performance data collected from Assam, India. The results of their study emphasised the significant impact of ongoing evaluation on determining final semester results, with the Random Forest (RF) classifier demonstrating superior accuracy compared to alternative approaches. Similarly, Almarabeh (2017) used the Weka software tool to assess the academic accomplishments of university students. This study compared the accuracies of different classifiers to draw meaningful comparisons.

To expand the focus of the study, Rao et al. (2017) used the capabilities of Weka to predict the consequences of student placement. The extensive study assessed the effectiveness of various machine learning algorithms and made comparisons across algorithms built in Weka and R Studio. The acquisition of such valuable insights has immense importance for higher educational institutions that want to enhance the quality of student instruction.

The study conducted the capabilities of Weka in the assessment of online admission systems for professional courses (Pant & Chowhan, 2023). The authors emphasise Weka's proficiency in classification, clustering, and association rule mining tasks. Hassan et al. (2022) used the Weka software to conduct feature selection analysis. Their study aimed to develop a prediction model that could effectively identify and predict digital addiction among student populations. The study shed light on the significance of Weka in determining characteristics that contribute to addiction across a range of features.

Pahuriray et al. (2022) used the Weka tool for sentiment analysis. The primary objective of their study was to analyse students' comments about flexible learning. The study results revealed the effectiveness of various machine learning algorithms in accurately predicting sentiment using Weka. Similarly, Al-Kindi and Al-Khanjari (2022) used the Weka software to compare the NB and RF classifiers in examining Moodle LMS course logfiles. The findings of their study revealed that the RF classifier outperformed the NB classifier.

Dol and Jawandhiya (2022) conducted an in-depth investigation of data mining methods, specifically highlighting the prominent role of Weka in the field of EDM. The researchers' comparative analysis of open-source and proprietary technologies highlighted the exceptional capability of Weka to effectively analyse datasets to identify patterns. The versatility of Weka in constructing classification models for practical situations was praised (Robu et al., 2023), who emphasised the simplicity of developing customised applications using the Weka API.

Weka was advocated to analyse feedback provided to hearing-impaired students at Politeknik Ibrahim Sultan (Arif et al., 2022). The researchers emphasised the importance of technical and vocational education and training in the fourth industrial revolution, highlighting the unexplored capabilities of students with disabilities. In their study, Ocaña-Fernández et al. (2023) integrated Weka with ensemble machine-learning methods to predict and assess the cognitive capacities of children diagnosed with autism.

The popularity of Weka in the field of EDM is well acknowledged. However, Weka's drawbacks include ARFF format requirement, end-user complexity, data preparation, limited scalability for big datasets, and Java environment needs (Robu et al., 2023).

c) Tableau

Tableau, a well-known data visualisation tool, has established a distinct position within the field of EDM (Tatnall, 2020). The software's user-friendly interface and exceptional capacity to integrate various datasets make it an important tool for educators and researchers seeking to extract significant insights from intricate educational data (KG et al., 2021).

Shafiq et al. (2022) conducted a comprehensive literature analysis focused on student retention, particularly emphasising the use of EDM and predictive analytics. The results of their study emphasised the effectiveness of visualisation technologies, such as Tableau, in effectively displaying student data. This enables educational institutions to make prudent options to enhance student retention. Similarly, Chakrabarty et al. (2023) praised the merits of Tableau in the field of interactive data analysis and visualisation.

Dermentzi et al. (2022) advocated for incorporating open data in academic settings, emphasising the significant contribution of Tableau in the analysis and visualisation of open data for students in higher education. The novel educational method presented by Murphy (2023) in teacher training revolves around the use of Tableau, with a strong focus on fostering student-centric learning. The model in question strongly resembles the dynamics seen in real-world educational settings, facilitating the development of a deep comprehension that surpasses traditional methodologies.

The seismic upheavals generated by the COVID-19 outbreak in computer science education in Nebraska using Tableau as a tool (Miller & Trainin, 2022). The investigation revealed significant changes in student demographics and patterns of interaction. Trogden et al. (2023) used Tableau as a tool to conduct a comprehensive and unbiased examination of undergraduate engagement. The primary objective of their study was to get insights into the details and consequences of student participation during an extended period.

Ahmad (2022) emphasised the effectiveness of digital dashboards, such as Tableau, in the context of inclusive education. Integrating these dashboards with various data sources has significantly transformed the data visualisation and reporting. Cadersaib et al. (2022) successfully incorporated Tableau into an adaptive framework within an academic setting. The integration successfully connected enterprise resource planning systems with big data analytics within the context of higher education. The programme aimed to align with the evolving demands of the business, promoting cooperation among many stakeholders and incorporating innovative teaching methods. The primary objective was to enhance student engagement and competency by providing immersive learning experiences.

In their comprehensive study, Lake et al. (2022) conducted a thorough assessment of the developmental path of accounting, with particular emphasis on integrating technical tools such as Tableau. Considering the forthcoming modifications to the 2024 CPA examination, educational establishments such as OSU are readjusting their academic programmes to conform to the evolving focus on technology.

Educational institutions employ Tableau, which is built for business intelligence. It may lack features for in-depth educational analysis due to its lack of EDM expertise (Chakrabarty et al., 2023). While Tableau's prowess in data visualization is acknowledged (Shafiq et al., 2022) and its application in education (Chakrabarty et al., 2023), its core functionality is tuned more towards business intelligence than EDM required in the current study. The software's limited capabilities in performing complex EDM processes necessary for precise student performance prediction led me to seek more specialised tools tailored for educational analysis.

d) KNIME

KNIME is well-recognised as a prominent open-source platform that excels in data analytics, reporting, and integration. The reason for its dominance in the field of EDM may be traced to its user-friendly graphical interface and ability to combine many data sources effectively (Acito, 2023).

Liu and Zeng (2023) emphasised the significant influence of DM in enhancing English teaching in vocational schools. The study conducted by the researchers used an enhanced version of the apriorist algorithm to investigate the intricate relationship between different courses and the instructional approaches applied. Significantly, the authors emphasised the crucial importance of KNIME in their prediction efforts, highlighting its important contributions to managing corporate data. Similarly, De Menezes et al. (2022) demonstrated the effectiveness of KNIME in developing customised data mining models for predicting academic performance among engineering students. The study achieved a notable accuracy rate of around 89.15%.

Cuéllar-Rojas et al. (2022) undertook a detailed text mining investigation with the KNIME platform, focusing on a more specialised application. Their endeavour aimed to identify clusters of topics within a large collection of abstracts focused on intelligent tutoring systems. Similarly, Hilmi et al. (2022) used the KNIME analytics platform to employ deep learning techniques to examine extensive online learning activities to enhance e-learning results.

Akhrif et al. (2022) advocated using the KNIME analytics platform due to its ability to integrate several data sources effectively, enabling a comprehensive study that enhances the predicting of student behaviours within intelligent learning environments. Velaj et al. (2022) comprehensively analysed a cutting-edge data science programme implemented at the University of Vienna. Designed specifically for those with a little grasp of computer science, this course, backed by KNIME, provided students with the ability to create visual workflows without coding. As a result, it significantly enhanced their computational skills.

Setiabudi and Santoso (2023) investigated the relationship between students' extracurricular participation and academic performances. Using data mining and clustering approaches, the researchers' results emphasised the need to maintain a careful equilibrium between involvement in extracurricular activities and academic pursuits. The use of the KNIME model for predicting student outcomes, considering several factors, highlights the diverse character of KNIME in analysing educational data.

While KNIME's capabilities in data analysis are well documented (De Menezes et al., 2022; Liu & Zeng, 2023), its suitability for current study in predicting student performance is limited.

KNIME’s low-code visual programming may not afford the granular control required for complex educational data modelling, and the potential for promoting uninformed practices and “black box” issues make it less desirable for the level of transparency and model interpretability (Ihrmark & Tyrkkö, 2023) in the current study.

e) Orange

Orange is segment based visual programming for DM, Artificial Intelligence (AI), and data examination, and with the help of orange tool researchers can visually represent to understand how certain relations can be depicted (Thange et al., 2021).

Rahman et al. (2022) undertook an investigation aimed at improving the effectiveness of programming teaching via the use of EDM techniques. The authors used an approach that utilised unsupervised algorithms, highlighting the importance of identifying significant patterns and rules within the data. In these attempts, tools such as Orange are considered essential since they provide understandable interfaces for activities such as classification, clustering and association rule mining.

In continuation of the discussion, Shafiq et al. (2022) conducted an in-depth investigation into the complex subject of student retention using advanced DM methods and predictive analytics. The researchers’ findings shed light on the potential of Orange as a crucial instrument for analysing student data, therefore providing educational institutions with valuable insights to enhance student retention. The notion of data fusion within the field of LA and DM was presented by Chango et al. (2022). The authors emphasised the importance of accuracy in data fusion approaches, particularly when integrating several multimodal learning analytics sources. The Orange platform emerges as a prospective cornerstone for effectively seamlessly integrating various diverse data sources within this setting.

Hwang (2023) used Orange 3, to thoroughly examine student comments about their experiences in maker education inside the metaverse. Similarly, Flores et al. (2022) used the Orange platform to develop prediction models focused on university student attrition. The study shed light on Orange's expertise in educational data sciences, with a specific focus on improving dropout prediction models. Yağcı (2022) demonstrated the efficacy of Orange as a predictive tool for precisely predicting students' academic performances. The integration of widgets into the tool's workflows streamlines the trajectory of data analysis for a wide range of users, including novices and experienced data enthusiasts.

Mike and Hazzan (2022) emphasised the capacity of Orange to facilitate the development of interdisciplinary collaborations among postgraduate academics in the fields of psychology and computer science. Similarly, Gresse von Wangenheim et al. (2021) advocated for using Orange's visual programming interface to enhance pedagogical practises in K-12 education. Sáiz-Manzanares et al. (2021) emphasised the significant role of Orange in the extraction and analysis of Moodle logs, which aids in the timely identification of students at risk in e-learning environments.

Orange is an impressive open-source data visualization, machine learning, and data mining toolset. Its visual programming front-end allows explorative qualitative data analysis and interactive visualization (Flogie & Krabonja, 2023). The adaptability and user-centric design of Orange make it an ideal choice for the current study in accurately predicting student performance. Its extensive toolset for data visualization and analysis (Rahman et al., 2022), and the capacity for seamless data fusion (Chango et al., 2022), align well with the multifaceted nature of educational datasets.

This section reviews five key EDM tools: RapidMiner, Weka, Tableau, KNIME, and Orange. While RapidMiner and Weka offer robust analytics but face scalability and usability limitations, Tableau excels in visualization but lacks specialized EDM functions. KNIME's low-code environment may limit modelling transparency. Consequently, Orange is selected for this study due to its open-source nature, user-friendly visual interface, seamless data fusion capabilities, and extensive toolkit for educational data mining and performance prediction, aligning best with the research requirements.

2.6 EDM Algorithms for Academic Performance Prediction

EDM has become a significant instrument in predicting student performance, providing educators with useful information to improve learning outcomes (Ben Said et al., 2023). The algorithms utilised in EDM play a crucial role by incorporating computational and statistical methodologies. EDM is an emerging discipline that uses diverse algorithms to analyse and make sense of extensive educational information (Yağcı, 2022). These algorithms facilitate discovering patterns, anticipating results, and extracting practical insights for researchers (Alalawi et al., 2023; Huynh-Cam et al., 2022). This section examines a selection of significant algorithms that have influenced the development of EDM, such as Naïve Bayes (NB), Decision Tree (DT), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs) (Batool et al., 2023), and these algorithms were used in current study.

a) Naïve Bayes (NB)

The NB algorithm has been more prominent as a probabilistic classifier within the domain of EDM. Considerable attention has been generated using this algorithm in predicting students' academic performance. One instance illustrating this phenomenon is the research (Yağcı, 2022), whereby several machines learning techniques, including NB, were used to predict the ultimate examination outcomes of undergraduate students based on their midterm scores as input variables. The predictive capabilities of the NB algorithm in EDM have been thoroughly documented and are supported by several studies (Alija et al., 2023; Nuarini et al., 2023; Sharma et al., 2023).

Sáiz-Manzanares et al. (2021) used several DM techniques, such as NB, to monitor and assess students' progress proficiently inside learning management systems. The study underscored the efficacy of NB in discerning students with a heightened likelihood of academic attrition. Ragab et al. (2021) suggested incorporating NB with ensemble methodologies to augment the precision of models used for predicting student performance. Based on the study results, integrative methodologies have led to a notable improvement in the precision of predictions.

Triayudi and Widyarto (2021) did a comparative study to evaluate the efficacy of J48 and NB algorithms within the domain of EDM to demonstrate the wide range of applications for NB. Their study findings underscored the superior precision of the NB classifier in predicting student performance compared to other classification algorithms. In the same manner, Perez and Perez (2021) used the NB classification algorithm to provide prognostications on programme completion rates at Bulacan State University. The results of their study revealed a significant level of accuracy, with a rate of 84%.

Nevertheless, it is important to acknowledge that while NB has shown exceptional performance in some scenarios, it has only sometimes emerged as the foremost contender. A comparative analysis was undertaken by Ariyanto and Chamidah (2021) to assess the performance of NB and Support Vector Machines (SVMs) in the context of sentiment analysis. The findings of the study indicate that SVMs had a significantly higher accuracy rate of 92.93%, whilst NB exhibited a comparatively lower accuracy rate of 79.86%. In a study by Garg et al. (2021), many classifiers, including SVMs, K-star (K*), Random Forest, and NB, were recognised as useful algorithms for predicting student performance, especially in scenarios with limited sample sizes. Xia and Yan (2021) extended the scope of applications by developing a framework for assessing music education. The researchers used the weighted NB algorithm and emphasised its potential in evaluating music performance.

Yahdin et al. (2021) have recently conducted a study that showcases an improvement in the precision of the NB algorithm. The use of the Relief-f algorithm for feature selection resulted in an enhancement of the accuracy rate from 73.43% to 74.38%. Rawal et al. (2023) and Vyasa et al. (2023) achieved accuracy rates of 72% and 70.3% respectively in their individual implementations of the NB classifier. Prince et al. (2023) and Saleh et al. (2023) used a fusion of NB and Collaborative Filtering algorithms in their individual study endeavours. Their objective was to augment the efficacy of delivering course and career recommendations to students while optimising learning strategies. Significantly, Saleh et al. (2023) had a remarkable accuracy percentage of 90.91% in which the study found the Decision Tree (DT) algorithm showed enhanced efficacy in evaluating student contentment during the COVID-19 pandemic, as compared to the NB algorithm (Kakish & Al-Eisawi, 2023).

NB algorithm's proven efficacy in predicting student performance (Sáiz-Manzanares et al., 2021; Yağcı, 2022), makes it a compelling choice for current study. Its capacity to process multifaceted educational data with notable accuracy aligns with the goal of developing precise predictive models for predicting academic performance in current study.

b) Decision Tree (DT)

DT has emerged as a prominent entity within the realm of EDM due to its inherent qualities of lucidity and comprehensibility. The ability to provide consistent results makes them a preferred choice for several researchers. Hung et al. (2020) established this phenomenon by using DT in combination with other DM algorithms to predict academic success in a blended learning course. Their study findings underscored the efficacy of DT in accurately identifying students vulnerable to academic underperformance throughout their educational journey.

Martín-García et al. (2019) emphasised the inherent flexibility and adaptability of DT. The researchers integrated the technology acceptance model with DT to investigate the intricate stages associated with the implementation of blended learning in higher education. The conducted study underscored the need to include blended learning as an independent variable inside prediction models for DT. Križanić (2020) delved into the realm of e-learning. The primary objective of the study was to use DT to analyse educational data derived from an e-course. The gathered data have yielded useful insights into the intricate patterns of student involvement with the e-learning platform, offering a view into their learning behaviours.

Wang (2022) examined the use of the C4.5 DT algorithm in the context of music teaching inside higher education institutions. The findings of this study revealed the potential of this algorithm to improve the quality of music education. Similarly, Yang et al. (2023) used the same algorithm to assess and enhance the competencies of research-oriented teachers with dual qualifications at vocational institutions. Matzavela & Alepis (2021) and Biehler & Fleischer (2021) have made significant contributions to the DT domain by researching intelligent mobile learning systems and pedagogical modules, respectively. The study has successfully tackled the gap between traditional educational approaches and data-driven insights.

Arcinas et al. (2021) and Palacios et al. (2021) have underscored the considerable relevance of DT in categorising and predicting student performance, thus accentuating its value. In the setting under consideration, Palacios et al. (2021) emphasised the better performance of DT in comparison to other machine learning methods. Duan et al. (2021) used a unique methodology including DT to evaluate the prevalence of smartphone addiction among young persons in China during the COVID-19 epidemic. Baashar et al. (2021) conducted a comprehensive analysis in their latest paper, wherein they explored several machine learning approaches used in the prediction of student outcomes. The authors placed particular attention on the significant contribution of DT in this situation. Ajibade et al. (2022) and Lee & Perret (2022) have enriched the academic discourse about DT. Ajibade et al. (2022) focused on investigating ensemble methodologies, while the research (Lee & Perret, 2022) centred on the design and implementation of a curriculum for AI education.

DT are widely acknowledged as a potent instrument in the field of EDM. This recognition is substantiated by several studies that have provided evidence of their efficacy across diverse educational contexts. The applications of DT include a broad spectrum, including domains such as predicting student performance, enhancing music pedagogy, offering intuitive visualization, easy interpretation, and handling both categorical and numerical data (Chen & Ding, 2023).

DT is particularly suited for the current study, given its exceptional ability to simplify complex decision-making processes in predicting student performance. The clarity with which DT models delineate the path from educational inputs to outcomes provides an intuitive means of understanding the underlying factors affecting student success. Their versatility across various educational contexts, as evidenced by studies from Hung et al. (2020) to Wang (2022), affirms their applicability in this study work, where accurate and interpretable results are paramount.

c) Artificial Neural Networks (ANNs)

ANNs have emerged as a crucial element in the domain of EDM, renowned for their exceptional capacity to discern complex data relationships and patterns. The diversity of their applications is seen in several situations, ranging from predicting disease outbreaks to comprehending student actions. Niazkar et al. (2020) emphasised the effectiveness of ANNs in predicting the occurrence of Covid-19 outbreaks, even when the connection to schooling seemed indirect. These qualities suggest their wider potential in educational contexts.

Expanding on the basis, Fernández et al. (2020) conducted a study in the field of interface mechanics, using ANNs to predict the constitutive characteristics of grain boundaries. The researchers' results provide light on the accuracy of ANNs in detecting intricate correlations within data, indicating a potentially fruitful direction for further exploration of student dynamics and learning trajectories. Coelho and Silveira (2017) conducted an extensive investigation of the increasing fascination with deep learning in EDM. The academic study emphasised the revolutionary potential of techniques such as ANNs in redefining educational paradigms.

In a more focused examination, Naim (2022) conducted a comprehensive investigation of ANNs, with a special emphasis on convolutional neural networks within the e-learning context. The study provided a comprehensive analysis of the effectiveness of ANNs in assessing student engagement in real-time, with a particular focus on the setting of online business education. In agreement with this viewpoint, Baashar et al. (2022) advocated for the versatility and precision of ANNs in predicting student academic performance, distinguishing them from traditional algorithms.

Sailer et al. (2023) explored the practical consequences of using ANNs to provide adaptive feedback to pre-service teachers to enhance their diagnostic reasoning abilities. Malik et al. (2023) demonstrated the effectiveness of ANNs in predicting student adaptation in the context of online entrepreneurship education. To validate their methodology, the researchers used the Kaggle educational dataset. Nguyen et al. (2023) provided a comprehensive overview of the revolutionary capabilities of ANNs in education. The authors also emphasised the ethical dilemmas arising from using ANNs, underscoring the urgent need for a worldwide agreement.

In addition to enhancing the ongoing discussion, Mittal et al. (2022) and Alam et al. (2022) have provided insightful viewpoints on ANNs. Mittal et al. (2022) conducted a study that centred on the utilisation of ANNs to manage stress in educational settings. In contrast, Alam et al. (2022) conducted to explore the integration of AI with firms in educational technology. Trivedi (2022) emphasised the effectiveness of ANNs in predicting student retention, highlighting its capacity to enhance educational results.

The dominance of ANNs in EDM cannot be disputed. The capacity to see and anticipate complex patterns has significant potential for the future of education. The increasing use of ANNs in education highlights the need for thorough study to address difficulties and enhance their integration in educational settings (Malik et al., 2023; Mittal et al., 2022; Sailer et al., 2023). ANNs are adept at modelling complex patterns, making them highly suitable for current study in accurately predicting student performance. Their deep learning capabilities allow for identifying subtle and non-linear relationships within educational data, offering insights that surpass traditional analysis methods. The adaptability of ANNs to various educational datasets and their proficiency in handling voluminous and complex data align with the objectives of the current study, promising to enhance the precision of academic outcome predictions.

d) Support Vector Machines (SVMs)

Support Vector Machines (SVMs), a supervised machine learning algorithm that can be used for classification or to solve regression problems. In practice, the SVMs algorithm is applied with the kernel that transforms an input data space into the required form (Yokkampon et al., 2021).

Sáiz-Manzanares et al. (2021) emphasised the crucial relevance of EDM approaches, particularly SVMs, in tracking student development inside learning management systems. The sentiment expressed by the user aligns with the findings presented in the study (Rodríguez-Arribas et al., 2021). The authors shed light on the possible applications of SVMs in active methodology, EDM, and learning analytics via their comprehensive mapping study. These insights give educators a refined perspective to examine student learning, enabling them to make well-informed judgements on intervention.

In their comprehensive study, Yi et al. (2022) conducted an in-depth investigation in the field of education, focusing on using SVMs to evaluate teaching effectiveness. The study conducted by the authors highlights the potential benefits of combining SVMs with multiple kernel learning and optimised parameters, providing a novel viewpoint on the subject matter. Simultaneously, researchers such as Munir et al. (2022) and Deepa et al. (2022) have emphasised the importance of SVMs in the realms of digital education and predictive analysis.

Hasib et al. (2022) conducted a comparison analysis to evaluate the effectiveness of several algorithms in predicting academic performances among secondary school students. Among the candidates under consideration, SVMs emerged as particularly noteworthy, with a notable accuracy rate of 96.89%. Khan et al. (2022) used SVMs in the context of vehicular ad-hoc networks to enhance the precision of route selection. This study demonstrates the potential for diversifying the applications of SVMs inside vehicular ad-hoc networks. Chen (2022) integrated SVMs with the grey wolf optimizer to predict the pedagogical quality for English majors. This study emphasises the significance of proactive learning attitudes.

Xiong and Lai (2023) and Pramudi et al. (2023) demonstrated the diverse applicability of SVMs in many disciplines, expanding their scope. Xiong and Lai (2023) investigated the assessment of teachers' skills using SVMs, whereas Pramudi et al. (2023) examined sentiment analysis in the context of data privacy rules. In their recent study, Hussain et al. (2022) made a notable contribution to the field by introducing Aspect2Labels, a system that utilises SVMs for sentiment analysis in the context of pedagogical feedback. This advancement significantly enhances the accuracy of sentiment analysis in this domain.

The indisputable dominance of SVMs in the field of EDM is evident. Their exceptional level of accuracy in both prediction and classification tests distinguishes them from others (Abdelkader et al., 2022; Phauk & Okazaki, 2021). SVMs are highly effective for current study on student performance prediction due to their superior classification capabilities and robustness in handling diverse and high-dimensional data. Their proven track record in EDM (Sáiz-Manzanares et al., 2021), provides a reliable method for discerning patterns that are critical for predicting academic success, validating their applicability and potential to enhance the accuracy of the predictive models of the current study.

2.7 EDM for Association Rules and the Use of DT Algorithm to Extract the Rules for High-Performance Students

This section uses the application of EDM through association rules and DT to elucidate the factors influencing high-performing students, blending complex data analysis with educational insights.

a) Using Association Rules in EDM

Association Rules in EDM are used to uncover hidden patterns in educational data (Dahdouh et al., 2020). These rules help in identifying relationships between different variables that impact student performance (Moubayed et al., 2018). The first step involves gathering comprehensive data on students' performance, including individual, family, school, and social dimensions and three compulsory subject performances. These data need to be cleaned and standardised for analysis. Key variables specific to each subject area are identified., like individual dimension about age (Arashpour et al., 2023), gender (Alemany-Arrebola et al., 2020), motivation level (Gumasing & Castro, 2023), self-efficacy (Fearnley et al., 2022), test anxiety (Jenal et al., 2022) and past performance (Alghamdi & Rahman, 2023); family dimension like parents' education level (Mehndiratta & Mehndiratta, 2023), family income (Alghamdi & Rahman, 2023), and parental involvement in students' learning (Leonard et al., 2022); school dimension like teacher's education (Early et al., 2007), title (Rivera Rodas, 2019), teaching methods (Assem et al.,2023) and self-efficacy (Wang, 2022), and social dimension like peer relationships (Fearnley et al., 2022), social support (Mishra, 2020), private tutoring (Yeung & Xia, 2023) and socioeconomic status (Issah et al., 2023). Using a tool like Orange, it can then mine the data to find frequent itemset. For instance, it might discover that students who spend a certain amount of time on private tutoring are more likely to perform well in Physics. The resulting rules, such as “High private tutoring participation → High performance in Physics”, help identify the characteristics of high-performing students in each subject.

b) Using DT Algorithm to Extract Rules

DT provide a visual and analytical model for decision making (Blockeel et al., 2023). They are especially useful for categorizing and predicting performance based on input variables (Blockeel et al., 2023). Using the same data, it can construct DT where each node represents a decision based on a variable (e.g., hours spent private tutoring, level of motivation) and each branch represents the outcome of that decision (Kumaran et al., 2022). The trees are trained with a portion of the dataset and tested with another portion to validate its accuracy (Ahmed et al., 2018). The goal is to minimize overfitting (where the model is too tailored to the training data) and ensure it generalizes well to new data, and once the trees are optimized, the rules can be extracted (Blockeel et al., 2023). For example, in a dataset used to predict student examination performance, the features include study time and class participation. The decision process is structured as follows: Root node: Study time > 2 hours; Left child node: Class participation > 70%; Leaf node: Predicts high performance. The rule derived is: If study time is more than 2 hours and class participation is more than 70%, the prediction is high performance. Here, study time and class participation correspond to the split conditions of the nodes. Through these EDM techniques, educators and researchers can gain a nuanced understanding of the factors contributing to student success in different academic areas (Chaka, 2021).

2.8 Students' Subject Options

This section delves into the evolving landscape of student subject options, contrasting traditional methods with contemporary data-driven approaches, and examining the impact of educational reforms and influences and insights on these options.

a) Method to Recommendations of Subject Options

Traditional methods in identifying characteristics of high-performing students typically involve observational studies, surveys, and statistical analyses. For example, Kuo et al. (2015) explored learning styles in EFL students, offering insights into how these styles impact performance in the English subject. Similarly, Huang et al. (2011) investigated the use of machine learning in educational settings, demonstrating how technology can assess and predict student learning performance. These studies provide a foundation for understanding the various factors that contribute to student success in academic settings.

The limitations of traditional methods in identifying characteristics of high-performing students are significant. Traditional statistical methods may fail to capture the complex interplay of factors affecting student performance. Studies like Huang et al. (2018) on learning style influence, and Biglan (1973) on subject matter characteristics, have highlighted these gaps. Huang et al.'s (2018) study underscored the challenges in correlating learning styles with academic performance accurately, while Biglan's study on disciplinary differences suggested a lack of specificity in understanding how subject matter influences student performance. These methods often overlook the interactions between individual, psychological, and environmental factors, leading to incomplete or generalised conclusions about student success.

In transitioning to current methodology, which utilises association rules and DT algorithm, they depart from traditional approaches by offering a more complex analysis of student performance characteristics. These data mining techniques enable the identification of subtle patterns and correlations in student data, which are often overlooked in simpler statistical models. Association rules can uncover hidden relationships between various student attributes and performance (Dahdouh et al., 2020), while DT offers a structured, hierarchical approach to classifying students based on a combination of factors (Ünal, F., 2020). This approach promises a more personalised and accurate understanding of what constitutes a high-performing student in different educational contexts.

The use of association rules and DT in analysing student performance offers several advantages. These methods excel in handling large datasets, allowing for the efficient processing and analysis of extensive educational data. They are particularly adept at uncovering hidden patterns and relationships within the data, which may not be apparent through traditional statistical methods (Dahdouh et al., 2020; Ünal, F., 2020). This leads to more personalised and accurate insights into student performance, facilitating tailored educational strategies. Studies like Holzinger (2019) on machine learning and Xiao & Hu (2019) on the use of AI in educational assessments provide evidence of the effectiveness of these advanced analytical techniques in the educational domain.

The objective of the current study is to identify the distinctive characteristics of high-performing students in each option subject. Subsequently, students may align their own characteristics with the recognised characteristics. If there is a match, then the subject is an appropriate choice. Therefore, association rules and DT algorithm will be very suitable for the above characteristic extraction and matching.

b) Subject Options in NCEE Reform

According to a study conducted by Huang et al. (2021), China underwent a significant educational reform in 2018, during which eight provinces and cities officially used a new approach to subject options in NCEE. This change represents a shift towards a system that better aligns with the goals of students and the general population, indicating a notable departure from the conventional and inflexible framework. The implementation of this reform has resulted in a discernible inclination among students and parents to choose for subjects seen as more financially advantageous, which may possibly result in an uneven distribution of subject instructors (Chen et al., 2022).

The reform saw a shift from the original “3+3” model to a “3+1+2” model, with a particular emphasis on enhancing the prominence of history and physics disciplines (Wang & Lei, 2018). The purpose of this transition is to address the challenges arising from the wide array of subject combinations and options, which have proven to be burdensome for students, schools, and educators. Students should make deliberate options about their academic disciplines, considering their interests, professional aspirations, and the level of competition associated with all subjects (Zhou, 2018). Additionally, students should also take into consideration their aptitudes.

The comprehensive NCEE score encompasses a wide range of academic areas, including Politics, History, Geography, Physics, Chemistry, and Biology, among others (Liang, 2021). The selection process is influenced by the standards established by educational institutions and the examinees’ knowledge and preferences. Furthermore, Zheng et al. (2019) assert that the implementation of educational reform has effectively abolished the previous dichotomy between arts and sciences in senior high schools. Consequently, students are now given the opportunity to choose optional subjects and undertake examinations that align more closely with their interests and aptitudes.

A comprehensive examination is undertaken to explore the significance of various academic disciplines and their implications for career advancement (Wang, 2021). Subsequently, these insights were disseminated throughout the students to enhance their decision-making process. Wang (2022) posits that the high school “3+1+2” examination approach entails the categorization of six minor subjects into two distinct groups - the first-choice subjects, namely Physics and History, and the second-choice subjects, which encompass Chemistry, Biology, Politics, and Geography. Under this approach, students are mandated to subject options from both categories. The newly implemented method has been specifically devised to effectively cater to the wide range of interests and aptitudes shown by students, therefore equipping them with the necessary skills and knowledge to excel in their future academic and professional pursuits.

Recent studies have focused on family, socioeconomic position, and personal preferences affecting academic subject options, but effective assistance for students' decision-making is lacking. Technical techniques like EDM may provide personalised recommendations via data analysis, but current literature generally ignores them. This gap indicates a lack of integrative educational theory-data science research that may provide more comprehensive insights. Retrospective analysis rather than proactive assistance for students' crucial options is also common. The present study fills these gaps by using the DT algorithm to uncover the characteristics of high-performing students in specific optional subjects. Students with matching characteristics will be recommended with the related optional subjects.

Subject options in education, particularly in critical disciplines like Chinese, Mathematics, and English, demand a data-informed approach to align with other optional subjects. Leveraging EDM, as Mkwazu and Yan (2020) advocate, current study dissected the NCEE's influence on students' subject options, aiming to predict performance and guide options that resonate with both personal interests and market demands (Yi et al., 2022). This synergy between student preference and predictive analytics form the cornerstone of current study's methodology.

2.9 Factors Affecting Student Academic Performance

In recent years, a significant amount of study has been conducted on many aspects that influence academic performance (Rajagukguk et al., 2023). Table 2.3 presents the comprehensive factors of the four distinct dimensions that impact student performance, as derived from an extensive review of relevant literature.

Table 2.3: Four Dimensions Affecting Students' Academic Performance Based on Chronological Order

Category	Individual Factors	Family Factors	School Factors	Social Factors	Authors

Table 2.3 continued

1			Teacher sensitivity, student-teacher rapport, pedagogical methods, and environmental challenges (low expectations and stigmatization)		Jacobson (2000)
2	Aggressive, impulsive, anxious behaviours and learning difficulties	Parental care and family finance	Educational expectations		Tu et al. (2009)
3	Attendance rates and gender differences			Cultural background	Moldabayev et al. (2013)
4			Collaborative methods, a culture of speaking English at school, regular and effective teaching, and leadership		Lekhetho (2013)
5	Reading fluency, comprehension, and vocabulary				Cimmiyotti (2013)
6	Attendance at science conferences/seminars		Teachers' salaries, the availability of instructional materials, and payment of teachers' allowance		Oredein & Awodun (2013)
7				Private tutoring	Zhan et al. (2013)
8	Task performance and task autonomy				Pulfrey et al. (2013)
9	Test anxiety, trait anxiety, and depression				Akinsola & Nwajei (2013)
10			Qualifications and teaching practices		Shannag et al. (2013)
11			Teacher attitudes	Shadow education, and market-driven tutorial centres	Chan & Bray (2014)
12	Academic success (high school CGPA)	Family background	Subject tracks	Tutoring	Berberoğlu & Tansel (2014)
13	Test-related anxiety and study skills		Test-taking training	Social dimension and social derogation	Putwain et al. (2016)

Table 2.3 continued

14	Students' desire for academic success and gender differences	Parental support and the number of books at home	Instructional quality	Socioeconomic background	Nilsen & Gustafsson (2016)
15			Organizational health dimensions		Zamora & Hernandez (2016)
16	Intrinsic motivation, self-efficacy, test anxiety, critical thinking and gender differences				Hamid & Singaram (2016)
17	Executive functions				Zorza et al. (2016)
18	Educational history	Family income of parental economic status	Inadequate facilities, ineffective rules and regulations, insufficient support, the quality of teacher-student relationships, and assessment methods	The desert climate and socio-cultural factors	Mamo et al. (2017)
19	Self-esteem and physical activity	Family relations and family affluence	School climate		Kleszczewska et al. (2018)
20		Parental involvement		Low-socioeconomic status	Duan et al. (2018)
21	Gender	Language spoken at home	Disciplinary climate, teacher-student relations, and resources like computers	The role of extracurriculars	Agasisti et al. (2018)
22			Teacher-student rapport, and school climate		Whittle et al. (2018)
23	Present pressure	Parental engagement and home strictness	Teacher attitudes towards students' studies	Future and job tensions	Saqib et al. (2018)
24	Student gender, unexplained absences, ethnicity, student age, learning techniques and study habits	Proper guidance from parents, parental educational and attainment	Learning facilities, communication skills		Mohamed et al. (2018)

Table 2.3 continued

25	High school GPA, self-confidence and independent activity focus	Institutional commitment		Brech & Burnett (2019)
26	Intrinsic and extrinsic motivation and learning behaviour			Tokan & Imakulata (2019)
27	Student's grade, attendance record and age	School subject	Student's neighbourhood and city	Fernandes et al. (2019)
28	Academic engagement	Digital readiness		Kim et al. (2019)
29	Learning motivation	Student enrolment and blended learning environment		Law et al. (2019)
30	Sleep quality, duration, and consistency			Okano et al. (2019)
31	Students' ability self-concepts, performance motivation and past success	Task values		Steinmayr et al. (2019)
32	Academic distraction	Technology usage		Feng et al. (2019)
33	Effective time management behaviours			Adams & Blair (2019)
34	Intrinsic motivation	Technology-enhanced learning type	Social media groups	Dunn & Kennedy (2019)
35	Cognitive presence levels (triggering event, exploration, integration)		Learner prominence measures (centrality in a learning community)	Galikyan & Admiraal (2019)
36	Executive functions (working memory, inhibition, cognitive flexibility, and planning), gender, age, physical fitness, motor skills and IQ			Cortés Pascual et al. (2019)
37	Psychological motivation, cognitive problem-solving and learning management	Interactions with instructors	Peer collaboration and community support	Lee et al. (2019)

Table 2.3 continued

38		Parental engagement			Lara & Saracostti (2019)
39				Social-emotional skills	Panayiotou et al. (2019)
40	Physical activity				Harveson et al. (2019)
41			Teacher quality (teacher's content knowledge and experience and self-efficacy beliefs)		Toropova et al. (2019)
42		Lower parental education levels	Guidance	Geographical isolation, socio-emotional support, socioeconomic status and barriers like long commutes	Echazarra & Radinger (2019)
43			School leadership, principal effectiveness, instructional leadership, teacher evaluation, professional development, school climate and teacher behaviours		Ozdemi (2019)
44	Emotion regulation	Lower-income backgrounds			Rozek et al. (2019)
45	Engagement in research activities	Lower family income and educated parents	Scholarships	Financial aid	Moreira et al. (2019)
46	Online learning activities and engagement, academic emotions (interest, enthusiasm, intrinsic motivation), and historical academic data		The teaching environment and style, and rapport between professors and students		Namoun & Alshanqiti (2020)
47	Self-efficacy, positive learning-related emotions, and metacognitive learning strategies				Hayat et al. (2020)
48	Past performance, emotional stability, and assessments (assignments and quiz marks)	Family characteristics (expenditure and income)	Student-instructor interaction		Waheed et al. (2020)

Table 2.3 continued

49	Students' background, experience, collaborations, interactions, and autonomy		Learning environment		Abuhassna et al. (2020)
50	Academic-related stress, anxiety about schoolwork, poor sleep, decreased motivation and engagement, mental health issues, and substance				Pascoe et al. (2020)
51	Health behaviours, and mental health issues	Parental education		Socioeconomic status	Alhadabi & Karpinski (2020)
52	Attitudes towards native speakers, and anxiety	Parental encouragement	Classroom environments, class sizes, and seating arrangements	Peer support	Getie (2020)
53	Learning motivation, and self-efficacy		Blended learning		Rafiola et al. (2020)
54	Historical grades and test scores, gender, and psychological attributes		E-learning activity	Socioeconomic status, environment	Alyahyan & Düştegör (2020)
55	Student engagement, academic self-efficacy and optimism		Collaborative learning	Social media	Ansari & Khan (2020)
56		Family resources (finances, academic guidance, parental involvement, language skills, and the educational level of parents)	Campus social organizations and faculty relationships	Social support	Mishra (2020)
57	Student motivation		Classroom interaction, course structure, instructor knowledge, and facilitation		Baber (2020)
58	Self-efficacy and anxiety levels, gender differences, and stress-induced anxiety				Alemany-Arrebola et al. (2020)

Table 2.3 continued

59	Meta-linguistic skills for reading and numerical skills for mathematics, cognitive abilities				Peng & Kievit (2020)
60	Self-efficacy (user confidence)		Technological facilities, and e-learning system quality, trust (security, privacy) and study environment	Cultural aspects (ICT literacy, social media integration)	Almaiah et al. (2020)
61			Identifying student cues for teaching adaptation, modelling effective learning strategies, and interactive methods		Littlejohn (2020)
62	Saccharomyces boulardii supplementation, and pre-examination stress				Karbownik et al. (2020)
63	Migrant status, ethnic minority status	Parental involvement	Student with teacher relationships	Higher socioeconomic status	Young (2020)
64			School tutoring	Family tutoring	Chui et al. (2020)
65	Interest/value in tasks, students' age, aptitude, and their attitude, dynamic of results, internal motivation		Classroom activities, and the reward-punishment, teaching methods	Relatedness from social rewards	Borah (2021)
66	Self-efficacy, gender		Learning environment		Talsma et al. (2021)
67	Diet, consistent sleep, physical activity, excessive screen time, and substance use				Sánchez-Hernando et al. (2021)
68	Effective study habits		Information/data management, digital creation skills, and digital independent learning		Limniou et al. (2021)
69	School grades, age, educational materials, and early semester study	Early engagement		Streaming video lectures	Cagliero et al. (2021)

Table 2.3 continued

70	Stress level, self-oriented perfectionism, and adaptability			Peer relationships	Miksza et al. (2021)
71	Attendance, study habits, physical resources, interest in study, regular study, hard work, dedication, self-confidence, and learning technique	Support by family			Islam & Tasnim (2021)
72	Class attendance, study hours, past academic results, and university admission test scores	Family income, and educated parents	Educational internet usage		Rahman (2021)
73		Parental income		Socioeconomic status	Mbogo et al. (2021)
74	Student expectations and prompt feedback		Instructor quality, and effective course design		Gopal et al. (2021)
75			Teacher burnout (teacher exhaustion and attrition)		Madigan& Kim (2021)
76	Motivation, positive emotions and negative emotions, cognition, and metacognition				Acosta-Gonzaga & Ramirez-Arellano (2021)
77	Intrinsic and extrinsic goal orientation, self-efficacy, control of learning beliefs, and test anxiety		Task value		Jenal et al. (2022)
78	Interactions, intrinsic goal orientation and meta-cognitive self-regulation	Private learning space			Lei & Lin (2022)
79	Perceived competence, coping strategies, and self-efficacy		Autonomy-supportive contexts, specific feedback, and growth mindset interventions		Liem (2022)
80	Cognitive load, self-efficacy, and long-term test performance		Item difficulty		Wu et al. (2022)

Table 2.3 continued

81	School marks, number of failed subjects, and time spent assimilating subject matters, personality traits, resilience, motivational processes, emotional intelligence, learning style, study strategies, self-efficacy, self-esteem, coping strategies, anxiety and stress control, study habits, creativity development, and personal satisfaction		Repeated courses	Socio-demographic variables, and social skills	Supervia et al. (2022)
82	The type and duration of physical activity				Shantakumar et al. (2022)
83		Home social capital, and parental engagement		Cultural capital, and shadow education (weekend tutoring)	Li & He (2022)
84		Parental warmth and parental involvement			Parmar & Nathans (2022)
85		Parental educational background, especially mothers' education		Higher socioeconomic status	Suningsih (2022)
86		Parental support, family size, level of parental income, and education level		Socioeconomic status	Leonard et al. (2022)
87	Learning efficacy, and positive emotions	Family cultural capital, family engagement, and parents' expectations			Luo et al. (2022)
88			Teacher emotional intelligence and engagement, and teacher self-efficacy		Wang (2022)
89	Self-efficacy		Interactions with instructors	Peer interaction	Fearnley et al. (2022)
90	The time spent on social media, health effects due to smartphone use	Family interaction	Educational impacts	Social interactions	Mathimagal et al. (2022)

Table 2.3 continued

91	Learning behaviours (the frequency of raised hands, resource utilization, engagement in discussions, and attendance)				Du et al. (2022)
92	Grades in various semesters	Family income, the number of siblings, whether the mother lives with the family, father's job, mother's job, accommodation type and location	School location		Alghamdi & Rahman (2023)
93	Technostress, cognitive appraisal, and coping strategies				Sharma & Gupta (2023)
94	Cognitive engagement (motivation and strategic learning), and emotional engagement				Khan et al. (2023)
95	Procrastination, and misconceptions		Innovative instructional strategies, and teachers' qualifications		Assem et al. (2023)
96	Prior academic performance, student behavioural traits, and psychological variables		School environment	Socioeconomic status	Issah et al. (2023)
97	Learning motivation				Gumasing & Castro (2023)
98	Engagement levels, past performance, age, and educational backgrounds				Arashpour et al. (2023)
99	The student's role		Formative and summative assessments, the role of the teacher, and the learning environment context	Changes in work restrictions, the adaptation challenges, and impact of the pandemic	Kasp & Venkatraman (2023)
100		Parental input, and family structure		Socioeconomic status	Guo et al. (2023)
101	Educational expectations of youths	Parental support for college education		Socioeconomic status	Yeung & Xia (2023)

Table 2.3 continued

102	Prior programming experience, and Class XII results	Parental education		Higher socioeconomic status, and extracurricular activities	Mehndiratta & Mehndiratta (2023)
103	Religiosity			Social provision, and social anxiety as ethnic minorities	Fakapulia et al. (2023)
104				Collaborative learning, social interaction, and social media usage	Nazeef & Ali (2024)
105			School bullying, Teacher-student relationships	Peer relationships	Deng (2024)
106			Principal leadership		Sudadio (2024a)
107	Procrastination, academic stress, academic self-efficacy, resilience, and motivation			Social anxiety	Munda & Tiwari (2024)
108		Education costs, parental participation	Teacher quality, curriculum relevance, facilities and resources	Community participation, education policies	Sudadio (2024b)

2.9.1 Individual Dimensions

Based on Table 2.3, the individual dimensions can be categorised into five aspects: psychological dispositions, learning behaviours and cognitive strategies, health-related dimensions, demographic characteristics, and past performance.

a) Psychological Dispositions

Table 2.4 shows the psychological dispositions derived from existing studies that are classified into five aspects, which include stress and test anxiety, self-esteem, emotions, motivation and self-efficacy.

Table 2.4: Aspects of Psychological Dispositions

Psychological Dispositions	
Stress and Test Anxiety	
Academic-related stress	Pascoe et al. (2020); Munda & Tiwari (2024)
Stress-induced anxiety	Alemany-Arrebola et al. (2020)
Pre-examination stress	Karbownik et al. (2020)
Stress level	Miksza et al. (2021)
Stress control	Supervia et al. (2022)
Test anxiety	Akinsola & Nwajei (2013); Hamid & Singaram (2016); Jenal et al. (2022)
Test-related anxiety	Putwain et al. (2016)
Self-esteem	
Self-esteem	Kleszczewska et al. (2018); Supervia et al. (2022)
Emotions	
Academic emotions	Namoun & Alshanqiti (2020)
Positive learning-related emotions	Hayat et al. (2020)
Positive emotions and negative emotions	Acosta-Gonzaga & Ramirez-Arellano (2021)
Positive emotions	Luo et al. (2022)
Motivation	
Intrinsic motivation	Hamid & Singaram (2016); Dunn & Kennedy (2019); Namoun & Alshanqiti (2020)
Intrinsic and extrinsic motivation	Tokan & Imakulata (2019)
Psychological motivation	Lee et al. (2019)
Performance motivation	Steinmayr et al. (2019)
Learning motivation	Law et al. (2019); Rafiola et al. (2020); Gumasing & Castro (2023)
Motivation	Pascoe et al. (2020); Baker (2020); Acosta-Gonzaga and Ramirez-Arellano (2021); Khan et al. (2023); Munda & Tiwari (2024)
Internal motivation	Borah (2021)
Motivational processes	Supervia et al. (2022)
Self-efficacy	
Self-efficacy	Hamid & Singaram (2016); Hayat et al. (2020); Ansari & Khan (2020); Almaiah et al. (2020); Alemany-Arrebola et al. (2020); Rafiola et al. (2020); Talsma et al. (2021); Supervia et al. (2022); Fearnley et al. (2022); Jenal et al. (2022); Liem (2022); Wu et al. (2022); Munda & Tiwari (2024)
Coping Strategies	
Coping strategies	Liem (2022); Supervia et al. (2022); Sharma & Gupta (2023)

According to Table 2.4, this category comprises factors related to psychological characters, including stress, anxiety, self-esteem, emotions, motivation, self-efficacy, and coping strategies. It reflects how students' internal psychological states and emotional management abilities can influence academic performance.

In a study examining the multifaceted impacts of stress on academic performance, Pascoe et al. (2020) identified a correlation between academic-related stress and reduced academic performance. This connection was further elaborated upon by Alemany-Arrebola et al. (2020), who observed a worsening of stress-induced anxiety because of the sudden shift to online learning frameworks during the pandemic, undermining students' confidence in their academic abilities. Furthermore, Karbownik et al. (2020) delineated a direct link between pre-examination stress and cognitive impairment, predicting a subsequent decline in both academic performance and psychosomatic health. Echoing these sentiments, Miksza et al. (2021) reported an inverse relationship between stress levels and subjective vitality among music students, implicating a detrimental impact on their educational engagement. Munda & Tiwari (2024) found that academic stress negatively impacts students' academic performance, with procrastination acting as a mediating factor between stress and success. Finally, Supervia et al. (2022) posited that stress control is pivotal, as it is interwoven with other personal determinants, such as self-efficacy and resilience, affecting student performance within the academic milieu.

Akinsola and Nwajei (2013) pointed out the harmful effects of test anxiety on academic performance. They advocated for multimodal interventions, including relaxation and cognitive restructuring, to enhance student performance. Putwain et al. (2016) contend that academic resilience should be cultivated to mitigate test-related anxiety. Jenal et al. (2022) distinguished between the adverse impacts of test anxiety and the beneficial effects of a belief in effortful success on academic performance. However, Hamid and Singaram (2016) showed that test anxiety and critical thinking were not clear or inversely related to performance.

Self-esteem emerged as a significant determinant of life satisfaction and academic performance (Kleszczewska et al., 2018). They advocated for enhancing self-esteem through physical activity to improve outcomes, notably for students from economically challenged backgrounds. Supervia et al. (2022) reinforced this view, recognizing that self-esteem, in conjunction with self-efficacy and resilience, had a nuanced effect on performance within the educational context. This shows that the effect of self-esteem on student academic performance receives further influence from other factors.

Emotions in academic, specifically interest and enthusiasm, significantly influenced student engagement and performance within online and blended learning realms (Namoun & Alshanqiti, 2020). Further, Hayat et al. (2020) discerned that positive learning-related emotions, spurred by self-efficacy, foster the application of metacognitive strategy, thus improving medical students' learning outcomes. Acosta-Gonzaga and Ramirez-Arellano (2021) identified a dichotomy wherein positive emotions aid motivation and strategy development, whereas negative emotions, like boredom and anxiety, undermine accomplishments. Finally, Luo et al. (2022) proposed that while positive emotions are conducive to online learning, the adverse effects of negative emotions can be mitigated by robust learning efficacy.

Intrinsic motivation, as revealed by Hamid and Singaram (2016), was strongly correlated with positive academic performance and well-being among medical students, with self-regulation playing a pivotal role in engagement. Tokan & Imakulata (2019) observed that intrinsic and extrinsic motivations and learning behaviour are significant determinants of performance. Dunn & Kennedy (2019) established that intrinsic motivation is a predictor of engagement and improved grades in technology-enhanced learning, whereas extrinsic motivation lacks a direct correlation with grades. The importance of psychological motivation for high-level academic activities in e-learning (Lee et al., 2019), while Steinmayr et al. (2019) demonstrated that ability self-concepts and task values hold more predictive power for academic success than goals and performance motivation. Acosta-Gonzaga & Ramirez-Arellano (2021) noted that motivation, coupled with emotions, cognition, and metacognition, significantly impacts learning, with positive emotions serving to enhance motivation. Supervia et al. (2022) underscored academic self-efficacy as a predictor of performance intertwined with motivational processes and satisfaction. The impact of ergonomic factors on learning motivation and, consequently, on academic performance during online classes was determined by Gumasing & Castro (2023). Cognitive and emotional engagements, especially motivation, were established as pivotal for academic success in primary schools (Khan et al., 2023).

Besides, intrinsic motivation, significantly influenced online engagement and was a key predictor of student performance (Namoun and Alshanqiti, 2020). Pascoe et al. (2020) demonstrated that stress, particularly when coupled with motivational decline due to sleep disruption, substantially affects performance. The crucial role of learning motivation in conjunction with blended learning methods for enhancing student performance (Rafiola et al., 2020), whereas Baber (2020) pinpointed motivation as vital for learning outcomes and satisfaction amidst the shift to online education during the pandemic. Borah (2021) emphasised that attitude, which directly impacts internal motivation and is shaped by task interest and teaching approaches, is central to academic success. The influence of learning motivation on social presence played a vital role in enhancing enrolment but did not directly influence learning performance in a blended learning setting (Law et al., 2019). Motivation played a crucial role in impacting student achievement, where procrastination linked to poor motivation results in reduced academic performance (Munda & Tiwari, 2024).

The strong linkage between self-efficacy, high motivation, and enhanced performance in medical students (Hamid & Singaram, 2016), underscores the criticality of belief in one's abilities. Hayat et al. (2020) found that self-efficacy bolsters learning-related emotions and metacognitive strategies, culminating in improved academic performance. The pandemic-induced anxiety diminished students' academic self-efficacy (Alemany-Arrebola et al., 2020), revealing the psychological foundations of self-belief. Almaiah et al. (2020) highlighted the direct influence of self-efficacy on engagement and success within e-learning contexts during the pandemic. Fearnley et al. (2022) concurred, recognizing self-efficacy as a significant predictor of academic performance and perceived learning, especially in online education contexts. Munda & Tiwari (2024) described that academic self-efficacy moderated the relationship between academic stress and procrastination, with high self-efficacy students actively delaying tasks but ultimately achieving better performance.

As posited by Ansari and Khan (2020) as pivotal for student engagement, self-efficacy directly affects academic performance, particularly evidenced by online lecture involvement and examination results. Jenal et al. (2022) noted that control beliefs, inherent to self-efficacy, foster a mindset where effort equates to success, thereby potentially boosting performance. Liem (2022) identified self-regulatory self-efficacy as a deterrent to perceived distractions and an enhancer of positive emotions, contributing to academic success. Wu et al. (2022) observed that students' flow states, indicative of immersive engagement, are amplified by high self-efficacy, thus improving performance. During the challenges of the COVID-19 pandemic, the role of self-efficacy in mobilizing performance resources (Talsma et al., 2021). Supervia et al. (2022) underscored self-efficacy as a critical performance indicator, interfacing with engagement and satisfaction and mediating resilience's impact on academic success. However, Rafiola et al. (2020) discerned that self-efficacy alone did not have a significant positive influence on student performance.

Coping strategies, when paired with perceived competence, enhanced academic performance and well-being (Liem, 2022). Resilience and coping strategies' significant impacted on academic success, particularly in adapting to social challenges (Supervia et al., 2022). Further, the importance of coping strategies in managing technostress in technology-enhanced learning environments, thereby affecting academic success, was corroborated by Sharma & Gupta (2023). These findings collectively emphasise the pivotal role of coping strategies in educational outcomes.

In the context of NCEE, individual psychological dispositions may affect student performance. The current study examines four elements, which include stress and test anxiety, motivation, and self-efficacy in all research questions.

b) Learning Behaviours and Cognitive Strategies

Table 2.5: Learning Behaviours and Cognitive Strategies

Learning Behaviours and Cognitive Strategies	
Habits	
Study habits	Mohamed et al. (2018); Limniou et al. (2021); Islam & Tasnim (2021); Supervia et al. (2022)
Metacognition	
Metacognition	Acosta-Gonzaga & Ramirez-Arellano (2021)
Woking Memory	
Woking memory	Cortés Pascual et al. (2019); Peng & Kievit (2020)

Table 2.5 shows study habits, metacognition, and executive functions such as working memory which are learning behaviours and cognitive strategies that affect academic performance.

Effective study habits, complemented by learning resources and parental involvement, are significantly linked to enhanced academic performance (Mohamed et al., 2018). Limniou et al. (2021) noted that resourcefulness and self-regulation, particularly in digital settings, are discipline-specific study habits conducive to student success. Furthermore, Islam & Tasnim (2021) posited that good study habits, hard work, dedication, and self-confidence are beneficial for academic performance, whereas lack of effort and interest are counterproductive. Supervia et al. (2022) recognised study habits as critical factors that, depending on their synergy within the educational sphere, can either foster or impede performance.

Metacognition, with its emphasis on students' self-regulation and cognitive awareness, has been identified as a pivotal element in academic performance (Acosta-Gonzaga & Ramirez-Arellano, 2021). Working memory is further recognised by Cortés Pascual et al. (2019) as a foundational factor, with its information management capabilities being vital for scholastic success. Peng and Kievit

(2020) expound on this, noting the role of working memory in facilitating essential cognitive functions and its resultant impact on learning.

Incorporating EDM into analysing learning behaviours and cognitive strategies reveals their crucial role in academic performance (Limniou et al., 2021). Current study leverages EDM to quantify the impact of these behaviours on performance in Chinese, Mathematics, and English, as part of China's NCEE. It aims to bridge the gap between cognitive strategies and subject options, aligning educational approaches with the cognitive assets of students, a connection underscored by the work of Peng and Kievit (2020).

c) Health-Related Dimensions

Table 2.6: Health-Related Dimensions

Health-Related Dimensions	
Activity	
Physical activity	Kleszczewska et al. (2018); Haeveson et al. (2019); Sánchez-Hernando et al. (2021); Shantakumar et al. (2022)
Sleep	
Sleep quality	Okano et al. (2019)
Poor sleep	Pascoe et al. (2020)
Consistent sleep	Sánchez-Hernando et al. (2021)
Healthy Habits	
Excessive screen time	Sánchez-Hernando et al. (2021)
Health behaviours	Alhadabi & Karpinski (2020)

Based on the Table 2.6, this category includes physical activity, sleep, and other health habits that can directly affect cognitive function and academic performance.

Physical activity had been shown to bolster self-esteem, subsequently enhancing life satisfaction and academic performance, particularly in socioeconomically disadvantaged families (Kleszczewska et al., 2018). Harveson et al. (2019) discovered that resistance physical activity acutely improves mathematics scores and cognitive function, underscoring its benefits to brain activity and scholastic tasks. Sánchez-Hernando et al. (2021) linked a lifestyle inclusive of regular physical activity with superior academic outcomes, positing that physical health underpins cognitive learning capacities. Additionally, regular physical activity like jogging and cycling were associated with better mental health and confidence, fostering improved academic performance (Shantakumar et al., 2022).

The crucial role of sleep quality and regularity for academic success was established by Okano et al. (2019), who proposed that sufficient, uninterrupted rest enhances cognitive functions in students. Pascoe et al. (2020) built on these findings, linking academic stress and poor sleep to a decline in scholastic performance, thus spotlighting anxiety's adverse effects on performance. Additionally, Sánchez-Hernando et al. (2021) reported that dietary habits, along with consistent sleep patterns and physical activity, have a direct bearing on educational outcomes, with balanced diets and adequate rest leading to improved grades.

The importance of healthy habits for academic performance is well-documented. Sánchez-Hernando et al. (2021) noted the negative impact of excessive screen time on academic performance. Alhadabi & Karpinski (2020) further established the link between health behaviours, mental health, and academic success.

Health-related dimensions such as physical activity and sleep quality are integral to cognitive function and academic performance (Kleszczewska et al., 2018; Okano et al., 2019). However, given the controlled environment of boarding schools in current study, these variables offer limited variability and thus will be outside the scope of current study on predicting performance in Chinese, Mathematics, and English using EDM. Instead, current study focus will be on leveraging EDM to analyse how other modifiable factors within the students' control may influence their subject options and performance in NCEE.

d) Demographic Characteristics

Table 2.7: Demographic Characteristics

Demographic Characteristics	
Gender	
Gender	Moldabayev et al. (2013); Hamid & Singaram (2016); Agasisti et al. (2018); Cortés Pascual et al. (2019); Talsma et al. (2021); Nilsen & Gustafsson (2016); Mohamed et al. (2018); Alyahyan & Düştegör (2020); Alemany-Arrebola et al. (2020)
Age	
Age	Mohamed et al. (2018); Cortés Pascual et al. (2019); Cagliero et al. (2021); Fernandes et al. (2019); Borah (2021); Arashpour et al. (2023)

Based on the Table 2.7, this category addresses factors such as gender and age.

Gender has been identified as a significant factor influencing academic performance. Moldabayev et al. (2013) observed that attendance rates impacted grades differently across genders, with male students initially showing a stronger correlation, which then shifted in later semesters. Although females generally exhibited higher motivation, this was not always reflected in significantly different academic performance compared to males, indicating a complex interrelation (Hamid & Singaram, 2016). Gender, along with school factors such as disciplinary climate and teacher-student relations, was influential in student performance, with supportive interactions and larger class sizes fostering resilience in a gender-differentiated manner (Agasisti et al., 2018). Gender differences were significant in the impact of executive functions, especially working memory, on academic success (Cortés Pascual et al., 2019). Gender was found to be a predictor of grades during the COVID-19 pandemic (Talsma et al., 2021).

Besides, Nilsen & Gustafsson (2016) highlighted the interplay of instructional quality, school climate, and gender in shaping motivation and performance. Interestingly, while females showed higher motivation, it didn’t necessarily translate to a significant difference in academic outcomes compared to males. Gender has emerged as a significant predictor of academic performance, as found by Mohamed et al. (2018), who suggested that societal and educational roles contribute to shaping academic outcomes. Alyahyan & Düştegör (2020) noted the inconsistent impact of demographics, including gender, across studies, hinting at a complex interaction with sociocultural factors. Additionally, Alemany-Arrebola et al. (2020) reported that during the COVID-19 pandemic, gender differences were pronounced, with women experiencing higher stress levels than men, negatively affecting their academic self-efficacy.

Age was identified as a significant predictor of academic performance, with maturity potentially enhancing learning through effective techniques and supportive home environments, as found by Mohamed et al. (2018). Older students often experienced lower success rates, possibly due to less favourable high school outcomes (Cagliero et al., 2021). Fernandes et al. (2019) highlighted age's role within the social environment affecting learning, with older students' demographic factors becoming more salient in their academic performances. Borah (2021) suggested that age's influence on performance was mediated by its interaction with teaching methods, learner aptitude, and attitudes. An inverse relationship between age and performance (Arashpour et al., 2023), with younger students outperforming older counterparts who, despite additional responsibilities, displayed resilience in academic improvement over time. However, Cortés Pascual et al. (2019) echoed this resilience, finding that age was not a significant predictor of performance, implying that cognitive abilities could outweigh the potential educational disadvantages of older age.

Demographic characteristics, specifically gender and age, will be integral to current EDM-based study on subject options in NCEE. Gender (Hamid & Singaram, 2016; Talsma et al., 2021), and age (Arashpour et al., 2023; Mohamed et al.,2018), are predictors of academic performance. Current study will incorporate these variables to understand their complex interplay with academic performance in Chinese, Mathematics, and English, and to derive insights that could inform tailored educational strategies within the context of recent NCEE reforms. Ethnic minority status and educational history, while influential, will not be central to current study analysis due to the homogeneity of the sample population.

In the context of NCEE, individual demographic characteristics may affect student performance. The current study examines two elements, which include age and gender in all research questions.

e) Past Performance

Table 2.8: Past Performance

Past Performance	
School CGPA	Berberoğlu & Tansel (2014); Brech & Burnett (2019);
Past successes	Steinmayr et al. (2019)
Historical grades and test scores	Alyahyan & Düştegör (2020)
Historical academic data	Namoun & Alshanqiti (2020)
Past performance	Waheed et al (2020); Arashpour et al (2023)
Past academic results	Rahman (2021)
School marks and failed subjects	Supervia et al. (2022)
Class XII results	Mehndiratta & Mehndiratta (2023)
Semesters grades	Alghamdi & Rahman (2023)
Previous academic performance	Issah et al. (2023)

Based on Table 2.8, school-CGPA, past successes, historical grades and test scores, historical academic data, past performance, past academic results, school marks and failed subjects, Class XII results, semesters grades, and previous academic performance are also very common predictors of student academic performance.

The influence of high school CGPA on performance was noted, with private tutoring providing a modest advantage, particularly in subjects like mathematics (Berberoğlu & Tansel, 2014). The significance of high school GPA as a harbinger of collegiate success was underscored, demonstrating a direct linkage between earlier and later academic performance (Brech & Burnett, 2019). Steinmayr et al. (2019) expanded the scope of predictors to include motivational constructs, indicating that students’ self-concepts and task values, shaped by past successes, influence subsequent performance. Alyahyan and Düştegör (2020) identified historical grades and test scores as primary indicators of academic success, while acknowledging the varied effects of demographics and environmental factors. Historical academic data was reported to significantly affect performance in online and blended learning contexts (Namoun & Alshanqiti, 2020). Past performance was identified as a robust predictor of future academic success, with early performances laying the groundwork for continued excellence (Waheed et al., 2020). Additionally, engagement levels, alongside previous success, were seen to shape educational outcomes, suggesting that past performance bolsters engagement and ongoing performances (Arashpour et al., 2023). Researchers conducted their studies by analysing various academic metrics, such as high school GPA, historical grades, test scores, and past performance data, alongside factors like private tutoring, motivational constructs, demographics, and student engagement, to identify key predictors of academic success across diverse learning contexts.

Furthermore, past academic results were found to be strong determinants of students’ GPA, with Rahman (2021) emphasizing their long-term impact on current success. Supervia et al. (2022) pinpointed school marks and failed subjects as reliable predictors of academic performance, suggesting that historical scholastic records underpin the development of students’ academic self-efficacy and resilience. The influence of Class XII results on performance in programming was examined using the apriorist algorithm to generate association rules (Mehndiratta & Mehndiratta, 2023). Alghamdi and Rahman (2023) identified grades across various semesters as significant indicators of performance, noting that family factors like income and parental involvement also play a role, thus indicating that past performance is influenced by both individual and familial factors. Confirming this trend, Issah et al. (2023) reported that previous academic performances are widely recognised as significant predictors of student learning outcomes. Therefore, past performance is an important influencing factor to consider in current study.

Past performance is a significant indicator for future success (Steinmayr et al., 2019; Waheed et al., 2020). In current EDM-based investigation into subject options for NCEE, historical academic performances is a key variable. By analysing past performance data, the study aims to predict outcomes in Chinese, Mathematics, and English, recognizing that prior success often sets a precedent for future academic endeavours, and tailoring educational strategies accordingly.

In summary, in current study, individual dimensions like age, gender, past performance, test anxiety, self-efficacy, and motivation levels are considered crucial for predicting academic performance in Chinese, Mathematics, and English, as supported by various research findings. Age and gender are recognised as significant predictors of academic performance. Studies show that gender can influence academic outcomes, with differences in motivation and performance observed between male and female students (Hamid & Singaram, 2016; Talsma et al., 2021). Age also plays a role, with older students displaying different learning patterns and achievements compared to their younger counterparts (Arashpour et al., 2023; Mohamed et al., 2018). Past academic performance, such as school CGPA and historical grades, is a robust predictor of future academic success (Berberoğlu & Tansel, 2014; Waheed et al., 2020). This historical data can provide valuable insights into a student's academic trajectory. Test anxiety adversely affects academic performance, as studies have shown a correlation between high levels of test anxiety and reduced academic achievements (Akinsola & Nwajei, 2013; Jenal et al., 2022). Self-efficacy, or a student's belief in their own academic abilities, is crucial for performance. Higher self-efficacy is often linked to better academic outcomes (Hamid & Singaram, 2016; Supervia et al., 2022). Motivation, both intrinsic and extrinsic, significantly impacts academic performance. Students with higher motivation levels tend to perform better academically (Supervia et al., 2022; Tokan & Imakulata, 2019). These dimensions are interrelated and provide a multifaceted view of student performance, essential for creating predictive models in EDM.

2.9.2 Family Dimensions

Based on Table 2.3, the family dimensions can be categorised into three aspects: demographic dimensions, parental involvement and support, and family environment.

a) Demographic Dimensions

Table 2.9: Demographic Dimensions

Demographic Dimensions	
Family Education Levels	
Parental education	Mohamed et al. (2018); Alhadabi & Karpinski (2020); Mehndiratta & Mehndiratta (2023)
Lower parental education levels	Echazarra & Radinger (2019)
Educated parents	Moreira et al (2019); Rahman (2021)
Educational level	Mishra (2020); Leonard et al. (2022)
Mothers' education	Suningsih (2022)
Family Income	
Parental income	Mbogo et al (2021)
Family income	Rahman (2021); Waheed et al. (2020); Alghamdi & Rahman, 2023
Levels of parental income	Leonard et al. (2022)
Family income of parental economic status	Mamo et al. (2017)
Source of family income	Mohamed et al. (2018)
Lower income	Rozek et al. (2019)
Lower family income	Moreira et al. (2019)
Family Structure	
Family structure	Guo et al. (2023)

Based on the Table 2.9, demographic dimensions include parental education levels, family income, and family structure.

Parental educational attainment was identified as a significant predictor of academic performance, with home-related aspects and study habits, influenced by parental education, positively correlating with student success (Mohamed et al., 2018). Parental education, as part of a broader set of socioeconomic factors, correlates with GPA and fosters non-cognitive skills like perseverance (Alhadabi & Karpinski, 2020). Parental education, particularly in programming subjects, was found to often promote better student outcomes (Mehndiratta & Mehndiratta, 2023). In rural contexts, lower parental education levels were shown to affect the degree of involvement in children's education and career aspirations, impacting academic performance (Echazarra & Radinger, 2019). Moreira et al. (2019) demonstrated that scholarships for study, potentially more accessible to students with educated parents, were associated with better academic performance in medical students. The education of both mother and father emerged as important factors that positively affect students' performance, with educated parents playing a supportive role (Rahman, 2021). Mishra (2020) noted that the educational level of parents affects the resources and guidance available to underrepresented students. Leonard et al. (2022) showed that higher parental education levels lead to better student support and success, while Suningsih (2022) found that parents' education, especially mothers', significantly motivates children in learning English as a foreign language.

Parental income was positively correlated with improved academic performance, suggesting enhanced resources boost exam success (Mbogo et al., 2021). Rahman (2021) demonstrated that not only family income but also class attendance, study hours, and educational internet usage were beneficial to students' CGPA, underscoring the importance of individual effort and supportive environments. Family income and emotional stability were emphasised as significant determinants of student performance (Waheed et al., 2020). Family income's role in predicting student success was underlined, with a call for increased parental involvement to further academic performance (Alghamdi & Rahman, 2023). Higher levels of parental income were linked to better academic outcomes, a result of stronger support and encouragement (Leonard et al., 2022).

Moreover, family income was noted by Mamo et al. (2017) to influence student performance through parental economic status and the consequent access to educational resources. However, Mohamed et al. (2018) found that once controlling for other factors, the primary source of family income did not significantly predict academic performance variations, pointing to the strong impact of learning techniques and study habits on academic success. Rozek et al. (2019) found that lower-income students often experience increased stress during high-stakes exams in STEM fields, potentially impairing their performance. Financial aid, particularly scholarships for research, was observed to enhance academic performance among medical students, suggesting it as a tool to counterbalance the effects of lower family income and promote educational equity (Moreira et al., 2019).

Guo et al. (2023) extensively analysed the impact of family structures on academic performance, concluding that intact family structures tend to support higher academic performance through access to resources and co-parenting. The absence of one or both parents, particularly in single-parent families, often results in lower academic performance, primarily due to reduced family socioeconomic status rather than the family structure itself. However, they found that family structure influences middle-school students' academic performance through socioeconomic status and parental input, with intact families fostering higher performance than those with absent parents.

The relationship between family demographic factors and academic performance is well-documented, with parental education and income levels being particularly influential (Alghamdi & Rahman, 2023; Leonard et al., 2022). In current study on subject options using EDM, it will explore how these factors affect students' performance in NCEE. Parental education level will be considered for its impact on the provision of educational resources and support, while family income will be examined for its role in accessing educational opportunities, both of which are crucial in shaping students' decisions and performance in key subjects.

In the context of NCEE, family demographic dimensions may affect student performance. The current study examines three elements, which include father's education level, mother's education level, and family income in all research questions.

b) Parental Involvement and Support

Table 2.10: Parental Involvement and Support

Parental Involvement and Support	
Parental Involvement	
Parental engagement	Saqib et al. (2018); Lara & Saracostti (2019); Li & He (2022)
Parental participation	Sudadio (2024b)
Family engagement	Luo et al. (2022)
Parental involvement	Duan et al. (2018); Young (2020); Parmar & Nathans (2022)
Parental care	Tu et al. (2009)
Parental encouragement	Getie (2020)
Early engagement	Cagliero et al. (2021)
Parental Support	
Parental support	Nilsen & Gustafsson (2016); Leonard et al. (2022); Yeung & Xia (2023)

Based on the Table 2.10, factors such as parental involvement and support are so important to affect student academic performance.

Parental engagement, such as teacher attitudes and home strictness, were found to significantly affect student academic performance (Saqib et al., 2018). Parental engagement was identified as instrumental in children's academic success, with higher involvement levels linked to improved performance (Lara & Saracostti, 2019). Parental engagement was also seen to overshadow financial contributions, offering a positive impact on academic outcomes, particularly in underprivileged families (Li & He, 2022). According to Sudadio (2024b), parental involvement influenced student motivation and performance by contributing to improved educational outcomes. Luo et al. (2022) expanded on this, noting the role of family cultural capital and expectations in fostering academic emotions and online learning engagement. For students from lower socioeconomic backgrounds, parental involvement was particularly beneficial, correlating with academic success and positive behaviour at school (Duan et al., 2018). Socioeconomic status was found to shape parental involvement, leading to proactive educational measures such as tutoring that enhance grades and exam results (Young, 2020). The effects of parental warmth and involvement on academic performance were found to vary by gender; in immigrant families, boys seemed to benefit more from such involvement (Parmar & Nathans, 2022).

In the absence of parental care, such as due to AIDS, children experienced lower academic marks and educational expectations (Tu et al., 2009). Parental encouragement was highlighted as beneficial for students learning English as a foreign language, indicating that home support builds positive attitudes and confidence crucial for language proficiency (Getie, 2020). Early engagement with educational materials and focused study efforts, rather than passive resource accumulation, were recommended for improving outcomes, emphasizing the importance of active involvement (Cagliero et al., 2021).

Parental support was highlighted as essential for academic success, contributing significantly to students' motivation and performance (Nilsen & Gustafsson, 2016). Leonard et al. (2022) observed that socioeconomic factors, including family size and parental education, significantly affected academic outcomes, with students from more affluent backgrounds often excelling due to stronger support systems. Parental support emerged as a critical factor influencing academic commitment and performance in middle school, with implications extending into adulthood (Yeung & Xia, 2023).

Parental involvement has emerged as a decisive factor in student academic performance, offering an essential layer of support that can substantially enhance educational outcomes (Lara & Saracostti, 2019; Leonard et al., 2022). Current study on the subject options of Chinese high school students will consider the depth of parental involvement as a predictive indicator for performance in key subjects. This aligns with current educational models that emphasise the role of family engagement in shaping students' academic trajectories and the importance of parental support in fostering a conducive learning environment.

In the context of NCEE, parental involvement and support may affect student performance. The current study examines parental involvement in students' learning in all research questions.

c) Family Environment

Table 2.11: Family Environment

Family Environment	
Language Environment	
Language spoken at home	Agasisti et al. (2018)
Parental language skills	Mishra (2020)
Private Learning Space	
Private learning space	Lei & Lin (2022)
Family Interactions	
Family interactions	Mathimagal et al. (2022)
Accommodation Type and Location	
Accommodation type and location	Alghamdi & Rahman (2023)

Based on the Table 2.11, family environment refers to the language environment at home, private learning space, family interactions, accommodation type and location.

Agasisti et al. (2018) identified the language spoken at home as a significant factor influencing student performance, with the linguistic environment playing a crucial role in academic outcomes. Mishra (2020) emphasised the importance of parental language skills, especially for underrepresented students, positing that home proficiency in the instructional language lays a foundation for success.

Lei & Lin (2022) discovered that a private learning space significantly boosts students' intentions to persist with online learning, suggesting that such spaces can improve concentration and academic outcomes. Family interactions were found to significantly affect student performance (Mathimagal et al., 2022). Furthermore, the type and location of student accommodation were identified as critical factors for academic performance, with the nature of living conditions, family income, and parental involvement all playing roles in academic success (Alghamdi & Rahman, 2023). Since residential students spend most of their time at school, the above three factors are excluded from current study.

The family environment, encompassing factors such as parental language proficiency and the quality of the home learning environment, plays a nuanced role in shaping educational outcomes (Agasisti et al., 2018; Lei & Lin, 2022). Current study will explore how these environmental dynamics influence students' subject options within the context of NCEE, recognizing the potential of such factors to either bolster or inhibit academic performance. However, the boarding school setting of current study means that specific environmental factors like private learning space and accommodation type are uniform and hence will not be included in the analysis.

In summary, the current study focuses on parental involvement, family income, and parents' level of education as critical family determinants of academic performance. Parental involvement has been consistently shown to have a significant impact on students' academic success. Studies indicate that higher levels of parental engagement are positively correlated with improved student performance, emphasizing the role of a supportive and involved home environment in fostering academic performance (Lara & Saracostti, 2019; Saqib et al., 2018). Family income also plays a crucial role, as a higher income typically provides access to better educational resources and support, directly influencing student performance. The positive correlation between family income and academic performance highlights the impact of socioeconomic status on educational opportunities (Alghamdi & Rahman, 2023; Mbogo et al., 2021). Furthermore, the level of education attained by parents is a significant predictor of student success. Higher parental education levels often result in more effective academic support and resources being available to students, thereby positively influencing their performance and aspirations (Mohamed et al., 2018; Rahman, 2021). Collectively, these dimensions shape the educational landscape and are pivotal in understanding and predicting academic performance, making them essential dimensions for analysis in EDM.

2.9.3 School Dimensions

Based on Table 2.3, the school dimensions can be categorised into two aspects: teacher dimensions, and school environment, resources and leadership.

a) Teacher Dimensions

Table 2.12: Teacher Dimensions

Teacher Dimensions	
Teacher Qualities	
Instructional quality	Nilsen & Gustafsson (2016)
Teacher’s content knowledge and experience	Toropova et al. (2019)
Instructor quality	Gopal et al. (2021)
Emotional intelligence and engagement	Wang (2022)
Self-efficacy	
Self-efficacy beliefs	Toropova et al. (2019)
Teacher self-efficacy	Wang (2022)
Teachers’ Qualifications	
Teachers’ qualifications	Shannag et al. (2013); Assem et al. (2023)
Instructional Practices	
Teaching practices	Shannag et al. (2013)
Pedagogical methods	Jacobson (2000)
Interactive methods	Littlejohn (2020)
Collaborative methods	Lekhetho (2013)
Teaching methods	Borah (2021)
Innovative instructional strategies	Assem et al. (2023)

Based on the Table 2.12, this category includes dimensions such as teacher qualities, self-efficacy, teachers’ qualifications, and instructional practices.

Instructional quality was identified by Nilsen & Gustafsson (2016) as essential for student motivation and performance. Toropova et al. (2019) demonstrated that a teacher's content knowledge and experience have a positive effect on mathematics performance. Gopal et al. (2021) noted that in online learning, instructor quality, student expectations, prompt feedback, and effective course design are critical for success. Additionally, Wang (2022) found that a teacher's emotional intelligence and engagement significantly predict student success.

Toropova et al. (2019) reported that while teacher self-efficacy beliefs are associated with perceived instructional quality, they do not significantly correlate with average math performance in classrooms. Contrarily, Wang (2022) highlighted that teacher self-efficacy, coupled with high engagement and emotional intelligence, enhances student outcomes, indicating the importance of teachers' belief in their capabilities.

Shannag et al. (2013) found that teachers' qualifications and student-centred teaching practices are key to student performance in science, with advanced degrees and certifications linked to better outcomes. Assem et al. (2023) also underscored that teachers' qualifications, particularly in subject-matter expertise and pedagogical skills, significantly influence student performance, with highly qualified teachers leading to greater performance. Teachers' qualifications do have a significant impact on the quality of teaching and learning, so there is necessarily included in current study.

Jacobson (2000) highlighted that pedagogical methods and positive student-teacher relationships significantly impact academic performance, with approaches that address individual needs and value diversity enhancing outcomes. The challenges of the transition to online teaching, particularly the lack of interactive methods, were discussed by Littlejohn (2020), noting their importance for developing proactive learning strategies. Collaborative methods and strong school leadership, coupled with a cultural shift towards English usage, were key for improving student performance in Lesotho (Lekhetho, 2013). Borah (2021) noted the influence of teaching methods on performance, mediated by students' attitudes shaped within the educational context. Innovative instructional strategies in physics, such as inquiry-based teaching, were found to boost engagement and comprehension, in contrast to traditional lectures (Assem et al., 2023).

Teacher qualities and instructional practices are instrumental in shaping academic success, as evidenced by the correlation between teacher expertise and student performance in subjects like mathematics (Toropova et al., 2019). These factors are critical in fostering an environment conducive to academic excellence and informed subject options within the framework of China's evolving educational landscape.

In the context of NCEE, teacher dimensions may affect student performance. The current study examines four elements, which include teacher's education, teacher's qualification, teaching methods, and teacher's self-efficacy in all research questions.

b) School Environment, Resources and Leadership

Table 2.13: School Environment and Resources

School Environment and Resources	
School Environment	
School climate	Kleszczewska et al. (2018); Whittle et al. (2018); Ozdemi (2019)
Blended learning environment	Law et al. (2019)
Study environment	Almaiah et al. (2020)
Learning environment	Abuhassna et al. (2020); Getie (2020)
School Resources	
Inadequate facilities	Mamo et al. (2017)
Technological facilities	Almaiah et al. (2020)
Facilities and resources	Sudadio (2024b)
School Leadership	
Leadership	Lekhetho (2013)
School leadership	Ozdemi (2019)

Table 2.13 continued

Principal leadership	Sudadio (2024a)

Based on the Table 2.13, this encompasses the overall school environment, resources available and leadership within the school.

Kleszczewska et al. (2018) found that a positive school climate, by promoting self-esteem through physical activity, especially in less affluent families, could enhance life satisfaction and improve academic outcomes. Whittle et al. (2018) elaborated that a school climate fostering active learning and strong student-teacher relationships boosts performance, despite potential time barriers. School climate also indirectly affects performance via school leadership, influencing teacher behaviours, which highlights the importance of effective leadership for academic success (Ozdemi, 2019). Law et al. (2019) found that the blended learning environment, through aspects like student enrolment and motivation, influences cognitive and social presence, which then indirectly impacts academic performance, whereas the teaching environment has a direct positive effect. The study environment, influenced by diverse factors like technology and resource readiness, plays a crucial role in academic performance (Almaiah et al., 2020). In online learning, the learning environment, shaped by student background, experience, and autonomy, is a significant determinant of academic success, with student satisfaction serving as a key mediator (Abuhassna et al., 2020). Getie (2020) pointed out that the classroom learning environment, alongside teacher attitudes and the availability of educational resources, affects language performance, with students' positive attitudes towards resources like textbooks reflecting favourably on certain educational context elements.

Mamo et al. (2017) found that institutional factors, including inadequate facilities and ineffective policies at Dire Dawa University, Ethiopia, were detrimental to female students' academic performance, signalling a pressing need for institutional reform. Almaiah et al. (2020) reported that during the COVID-19 pandemic, students' e-learning performance was positively influenced by technological facilities readiness and system quality, yet trust issues presented a negative impact. According to Sudadio (2024b), school facilities and resources enhanced student learning experiences and outcomes by providing adequate physical and technological support.

In Lesotho, Lekhetho (2013) found that strong leadership, alongside collaborative teacher-student efforts, and English language immersion, directly improved student performance on school-leaving examinations, underscoring the significance of leadership in academic success. Ozdemi (2019) also emphasised the crucial role of school leadership, particularly the indirect influence of principals through instructional leadership and teacher development, in enhancing student academic outcomes. Sudadio (2024a) argued that principal leadership was crucial for enhancing student performance, emphasizing its transformative nature and character changed within individuals. The research participant’s school leadership changes more frequently and is therefore not considered.

In summary, this current study considers teacher’s education, teacher’s qualification, teaching methods, and teacher’s self-efficacy as a pivotal school role in student academic performance. Teacher’s education and qualification directly impact the quality of instruction and student performance. Studies have shown that teachers with higher qualifications and specialised training facilitate better learning environments and student’s performance (Assem et al., 2023; Shannag et al., 2013). Teaching methods are equally crucial, as innovative, and student-centric approaches like collaborative and interactive methods have been linked to enhanced student engagement and performance (Borah, 2021; Jacobson, 2000). Lastly, teacher’s self-efficacy influences their teaching effectiveness and engagement with students, which in turn affects student learning performance (Toropova et al., 2019; Wang, 2022). These factors collectively contribute to a dynamic and effective educational environment, making them essential considerations for EDM in predicting academic performance.

2.9.4 Social Dimensions

Based on Table 2.3, the social dimensions can be categorised into three aspects: cultural influences and socioeconomic status, peer relationships and extracurricular activities, and social interactions and tutoring.

a) Cultural influences and Socioeconomic Status

Table 2.14: Cultural Influences and Socioeconomic Status

Cultural Influences and Socioeconomic Status	
Cultural Influences	
Cultural background	Moldabayev et al. (2013)
Socio-cultural factors	Mamo et al. (2017)
Cultural aspects	Almaiah et al. (2020)
Socioeconomic Status	
SES	Alhadabi & Karpinski (2020); Alyahyan & Düştegör (2020); Mbogo et al. (2021); Issah et al. (2023)
Low-SES	Duan et al. (2018)
Higher SES	Young (2020); Suningsih (2022); Mehndiratta & Mehndiratta (2023)

Based on the Table 2.14, factors such as cultural background and socioeconomic status (SES) play a pivotal role in shaping academic performance.

The cultural background was found to significantly affect student performance, as Moldabayev et al. (2013) highlighted its role in academic outcomes. Mamo et al. (2017) identified socio-cultural factors at Dire Dawa University as influential on female students' academic performance, with the cultural context affecting the availability of learning materials and services, thereby impacting success. Cultural aspects were also observed to enhance engagement and success in e-learning during COVID-19, with cultural initiatives improving student satisfaction and confidence (Almaiah et al., 2020).

Alhadabi & Karpinski (2020) observed that health behaviours and mental health, along with SES, correlate with GPA, emphasizing the substantial impact of non-cognitive factors such as grit and self-efficacy on academic performance, highlighting the importance of an educational environment that nurtures these qualities. Alyahyan & Düştegör (2020) noted the varied impacts of SES on academic outcomes, with its effects modulated by students' prior performances and environmental context. While acknowledging SES as a significant determinant of Mathematics and English performance, Mbogo et al. (2021) contested the notion that SES alone predicts academic success. Prior academic success, demographic traits, and SES as significant influencers of academic performance (Issah et al., 2023).

The benefits of parental involvement, especially for students from low-SES backgrounds, as it correlates with academic success (Duan et al., 2018). Young (2020) revealed that parents with higher SES are more likely to participate in their children's education, a factor that correlates with improved grades and performance on exams. Guardians of higher SES are active in supporting their children's English language learning, a commitment that enhances academic performance (Suningsih, 2022). Improved performance in programming subjects were linked to higher SES, which encompasses parental education, they noted SES's influence on prior experience and performance (Mehndiratta & Mehndiratta,2023).

In the context of NCEE, socioeconomic status (SES) may affect student performance. The current study examines SES in all research questions.

b) Peer Relationships and Extracurricular Activities

Table 2.15: Peer Relationships and Extracurricular Activities

Peer Relationships and Extracurricular Activities
Peer Relationships

Table 2.15 continued

Peer collaboration	Lee et al. (2019)
Peer support	Getie (2020)
Peer relationships	Miksza et al. (2021); Deng (2024)
Peer interactions	Fearnley et al. (2022)
Extracurricular Activities	
The role of extracurriculars	Agasisti et al. (2018)
Extracurricular activities	Mehndiratta & Mehndiratta (2023)

Based on the Table 2.15, this encompasses the impact of peer support, and extracurricular activities.

Lee et al. (2019) emphasised the importance of peer collaboration in e-learning for cognitive problem-solving and motivation. The significance of peer support in enhancing academic outcomes, who observed its positive effect on attitudes toward learning and interactions with native speakers (Getie, 2020). Deng (2024) stated that peer relationships can positively or negatively impact students' academic performance, with acceptance enhancing academic development and rejection, including bullying, negatively affecting academic performance. Miksza et al. (2021) identified the quality of peer relationships as essential for students' vitality and engagement in music education, suggesting that while stress diminishes vitality, strong peer bonds can counteract this effect. Conversely, Fearnley et al. (2022) noted that peer interactions, unlike those with educational content and instructors, had a lesser effect on satisfaction and academic outcomes.

Agasisti et al. (2018) recognised the role of extracurriculars in creating an environment favourable to academic success, intertwined with parental support, socio-economic status, and educational aspirations that together influence academic commitment and performance. Mehndiratta & Mehndiratta (2023) also found that extracurricular activities enhance student performance in programming, linking such involvement to socio-economic advantage and higher parental education, which afford students enriched experiential learning and support networks.

In the context of NCEE, peer relationships may affect student performance. The current study examines it in all research questions.

c) Social Interactions and Tutoring

Table 2.16: Social Interactions and Tutoring

Social Interactions and Tutoring	
Social Interactions	
Social media groups	Dunn & Kennedy (2019)
Social media integration	Almaiah et al. (2020)
Social media	Ansari & Khan (2020)
Community support	Lee et al. (2019)
Socio-emotional support	Echazarra & Radinger (2019)
Social support	Mishra (2020)
Social interaction and social media usage	Nazeef & Ali (2024)
Tutoring	
Private tutoring	Zhan et al. (2013)
Tutoring	Berberoğlu & Tansel (2014); Chui et al. (2020)
Weekend tutoring	Li & He (2022)

Based on the Table 2.16, Dimensions including social interactions and tutoring are vital factors that affecting student academic performance by many scholars.

Dunn and Kennedy (2019) found that student performance in a technology enhanced learning environment is positively influenced by active engagement with social media groups, suggesting a link between such interaction and better grades. Almaiah et al. (2020) indicated that e-learning is enhanced by factors including social media integration, which boosts engagement and success. Moreover, social media facilitates collaborative learning, significantly improving academic performance by fostering creativity and global learning connections (Ansari and Khan, 2020).

Community support and peer collaboration are also key social factors that impact student performance, fostering a sense of belonging and aiding problem-solving in e-learning environments (Lee et al., 2019). In rural settings, socio-emotional support plays a critical role in affecting student performance, academic participation and completion rates, as well as students' self-esteem and career aspirations (Echazarra & Radinger, 2019). Mishra (2020) highlighted that social support significantly influences the academic outcomes of underrepresented students, with community resources particularly beneficial in contexts lacking social capital. Nazwwf et al. (2024) described how social interaction with peers and teachers, as well as social media usage, positively impact students' academic performance through collaborative learning.

Private tutoring in one-on-one and small-group settings has been found to be effective for exam preparation and improving learning strategies in Hong Kong (Zhan et al., 2013). The impact of such tutoring, particularly in subjects like mathematics and the Turkish language, is minor compared to other factors such as socioeconomic status, and it may contribute to social inequalities (Berberoğlu & Tansel, 2014). An algorithm proposed by Chui et al. (2020) suggests that a holistic approach to tutoring can positively affect student performance. Weekend tutoring and parental engagement are especially beneficial for academic outcomes in less advantaged families, while tutoring during workdays may have a negative impact (Li & He, 2022). The above scholars hold two different perspectives, so the inclusion of private tutoring in current study is necessary.

Social support emerges as a crucial element in bolstering students' academic journey, providing a foundation for engagement and success that extends beyond the classroom (Mishra, 2020). Such support, especially in underrepresented communities, can significantly mitigate the challenges posed by socioeconomic disparities. In contrast, private tutoring, while beneficial, presents a dichotomy in its impact; it can enhance learning and exam preparation yet potentially exacerbate social inequalities (Berberoğlu & Tansel, 2014). Current study aims to dissect the influence of these social factors within NCEE framework, employing EDM to discern their roles in shaping students' academic decisions and performance in core subjects. This approach seeks to offer actionable insights that address the multifaceted nature of educational attainment.

In the context of NCEE, social interactions and tutoring may affect student performance. The current study examines two elements, which include social support and private tutoring in all research questions.

In summary, this current study considers social dimensions like private tutoring, social support, peer relationships, and socioeconomic status (SES) due to their significant impact on academic performance. Private tutoring has been shown to enhance learning strategies and exam preparation, although its effectiveness varies (Berberoğlu & Tansel, 2014; Zhan et al., 2013). Social support, including community and socio-emotional support, is crucial for underrepresented students, positively influencing academic performance (Echazarra & Radinger, 2019; Lee et al., 2019; Mishra, 2020). Peer relationships are identified as key to student vitality, engagement, and success, especially in collaborative learning environments (Lee et al., 2019; Miksza et al., 2021). Lastly, SES is a significant determinant of academic performance, influencing access to educational resources and overall learning environment (Alhadabi & Karpinski, 2020; Alyahyan & Düştegör, 2020; Mbogo et al., 2021). These dimensions collectively provide a comprehensive understanding of the diverse influences on student performance, making them vital for EDM in current study.

2.10 Theoretical Basis of the Study

Developed by Urie Bronfenbrenner in 1979, the Ecological Systems Theory (EST) systematically described the interactions between individuals and their environments (Bronfenbrenner, 1981), which emphasizes the importance of system-related factors and the interaction of biological factors and the contexts in which people develop (Rosa & Tudge, 2013). Significantly, the theory encompassed multiple levels of environmental systems: the Microsystem, Mesosystem, Ecosystem, Macrosystem, and Chronosystem. Figure 2.1 depicts the five systems posited by Bronfenbrenner's EST. These systems extended beyond immediate environments like classrooms and homes to broader social structures, such as educational policies and socio-cultural contexts, which, although not directly involving individuals, exerted a significant influence (Chong et al., 2023).

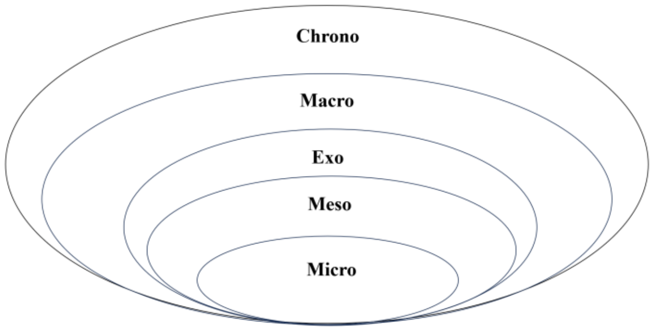

Figure 2.1: Bronfenbrenner's EST (Yang & Sanborn, 2021)

The following describes the five systems based on Guy-Evans (2024). Microsystem is the foundational level of Bronfenbrenner's model. It includes a child's immediate environment and relationships, like family and teachers, that have a direct, bidirectional impact on the child's development. For example, supportive parents can enhance cognitive skills, whereas negative peer interactions can affect self-esteem. The mesosystem represents the interactions between various microsystems in a child's life, influencing each other. Positive communications between parents and teachers promote consistency in the child's environments, while conflicts can lead to negative outcomes such as stress or poor academic performance. The next system is the exosystem. This level encompasses broader social systems that indirectly affect the child through their impact on the microsystem. For example, a parent's work conditions, or local school board decisions can alter the child's immediate environment and available resources.The macrosystem includes cultural norms and societal values that shape the conditions within which children develop. Cultural beliefs about gender roles and socioeconomic conditions are examples that profoundly influence a child's growth and worldview. Finally, the chronosystem involves the dimension of time, reflecting the personal and societal changes that affect a child throughout their life. Events such as parental divorce or societal shifts such as economic downturns shape the child's development over time.

As academic performance is an indicator of student learning and learning relates to an individual's behaviour, the current study employs this theory to encapsulate the numerous factors that affect academic performance at different system levels. Figure 2.2 depicts the distribution of factors derived in Section 2.9 across the five levels.

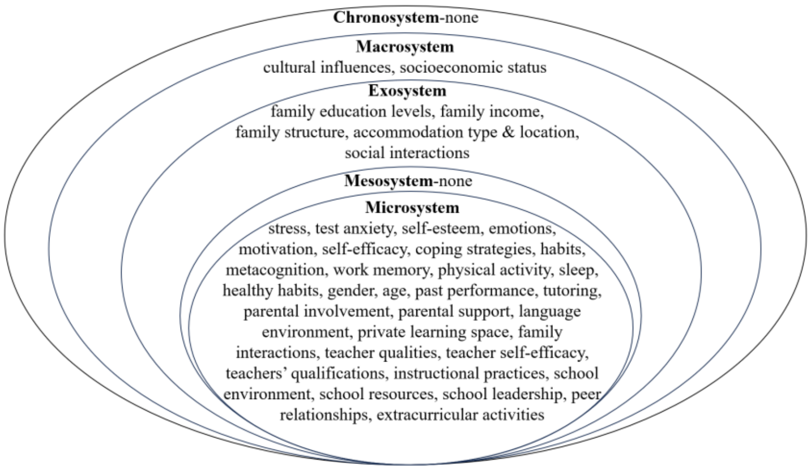

Figure 2.2: Framework Diagram of Influencing Factors Based on EST

The microsystem encompasses the individual dimension factors (Tables 2.4, 2.5, 2.6, 2.7, 2.8, 2.16 tutoring), family dimension factors (Tables 2.10, 2.11), teacher dimension factors (Table 2.12), and school dimension factors (Tables 2.13, 2.15). Section 2.9 currently lacks factors classified under the mesosystem. The exosystem includes the family dimension factors (Tables 2.9, 2.11 accommodation type & location) and the social dimension factors (Table 2.16). The social dimension factors from Table 2.14 are classified as the macrosystem, and there is no factor classified under the chronosystem.

2.11 Chapter Summary

This chapter mainly reviews the literature related to CEE, NCEE, DM, EDM, students' subject options, affecting factors, and the theoretical basis of this study, and analyses the research gaps, which lays a good foundation for this study. Research methodology is explored in the next chapter.

CHAPTER 3

METHODOLOGY

3.1 Introduction

The present chapter explains the research methodology of the study. This includes the research setting, research design, data collection processes, and research ethics.

3.2 Research Setting

The study was carried out in a prestigious Chinese high school, known as High School F. High school F, situated in a vibrant metropolitan setting, serves as a typical example of Chinese high schools that aim to equip students for NCEE. The educational institution comprises students with various academic ambitions and diverse educational backgrounds who are from different cities and schools of different nature.

3.3 Population and Sample

1289 students who take NCEE in 2025 and 103 teachers corresponding to these classes constitute the total population of this study, with 1127 students and 88 teachers serving as the final sample.

As indicated in Chapter 2, Table 2.4 - 2.16 shows the factors like age, gender, past performance, test anxiety, self-efficacy, motivation levels, parental involvement, family income, parents' level of education, teacher education, teacher qualifications, teaching methods, teacher self-efficacy, private tutoring, social support, peer relationships, and socioeconomic status (SES) for each dimension that are examined to predict students' performance in Chinese, English and Mathematics and to uncover the characteristics of high-performing students in examined optional subjects. The tables also show the means to collect the data for the respective factors.

To have a thorough comprehension of the academic paths of the students, the study used a multifaceted data gathering technique. The multifaceted data gathering technique was implemented by combining qualitative teacher feedback with quantitative academic records from a comprehensive school database, followed by systematic computerized post-processing for longitudinal analysis. According to Zhuoma (2023), the feedback given by teachers offers a distinct viewpoint on the performance of students, emphasising both their areas of proficiency and regions that have room for growth. The academic records of students at high school F are stored in a comprehensive database, which includes historical performance indicators, subject options, and other pertinent information. The database functioned as a valuable repository of data, enabling a comprehensive examination of students' educational progress over an extended period. After the collection of data, it underwent thorough post-processing inside a computerised setting.

Research design is the conceptual blueprint that systematically outlines the methods and procedures for collecting, measuring, and analysing data to address a research problem. It plays a crucial role in guiding the whole investigation, providing a comprehensive plan that delineates the necessary processes and techniques to be used to accomplish the research goals. This research design incorporates a systematic methodology for comprehending and predicting student performance within the context of NCEE, drawing upon EDM design put forward by Malini and Kalpana (2021), see Figure 3.1.

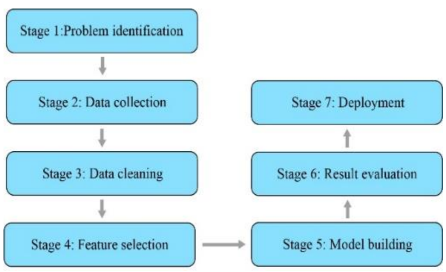

Figure 3.1: Research Design

3.3.1 Problem Identification

As explained in the problem statement section in the first chapter, there is a need to examine factors affecting students' performance in NCEE compulsory subjects using EDM. The application of EDM techniques in predicting student performance is relevant due to the substantial data volume within educational databases. Chapter 2 unveils various factors from the individual, family, school and social dimensions that can affect academic performance, and this study will examine how some of these factors can predict students' performance in the NCEE compulsory subjects, which include Chinese, English and Mathematics. Table 2.4 - 2.16 show the examined factors for each dimension.

To the researcher's knowledge, there is no relevant scholars have given scientific advice on the subject options for NCEE. Hence, this study investigates the characteristics of high-performing students in Physics, Chemistry, Biology, Politics, History, and Geography. Identified characteristics will inform the recommendations for subject options. Similar factors from the individual, family, school, and social dimensions are employed in extracting these characteristics.

3.3.2 Data Collection

Collecting data is a crucial step in research as it ensures that findings and interpretations rely on real evidence (Cimini et al., 2020). Initially, the surveys were conducted on 1289 students and 103 teachers. These 1,289 teachers are the instructors for these 103 students. However, some students' information was incomplete and some students and teachers refused to participate in the surveys. Therefore, a portion of the data was excluded. Eventually, this study collected data from 1127 students and 88 teachers for further analysis.

a) Survey Instrument

The survey utilized in this study consists of 122 questions and is directly adopted from existing surveys without changes. All survey items are presented in English, as the participating high school students have demonstrated sufficient proficiency in the language to ensure accurate comprehension and valid responses. The student questionnaire collects demographic, familial, and social factors, while the teacher questionnaire gathers professional background and teaching methodology data.

Factors such as age, gender, past performance are obtained directly from the school's existing database. Factors such as father's education level, mother's education level, family income, parental involvement in students' learning, peer relationships, social support, and private tutoring will use basic information through student questionnaire (see Appendix A). Teacher's education level, qualification, teachers' self-efficacy, will be conducted using teacher questionnaire (see Appendix B). Among all the factors, high school F currently has two teaching methods for teachers to choose namely traditional teaching method and smart classroom teaching methods. Socio-economic status is reflected through family income and parental education. There are a total of approximately 122 online questionnaires were be sent to respondents at high school F by the platform of Wenjuanxing for students and teachers (https://www.wjx.cn/). The unmodified instruments are presented in Table 3.1.

Table 3.1: The Unmodified Instruments

Category	Survey Factors	Instruments	Author (s)	Reference
1	Parental involvement in student's learning	Questionnaire	Akello, G. R. (2020)	Influence of Parents' Involvement in Education on Their Children's Performance at the Kenya Certificate of Primary Education in Nyakach Sub-County, Kisumu, Kenya
2	Motivation	4-point scale	Stover et al. (2012)	Academic Motivation Scale: adaptation and psychometric analyses for high school and college students. Psychology research and behaviour management, 71-83.
3	Self-Efficacy	Academic scale-2006	Gafoor & Ashraf. (2016)	Academic Self-Efficacy Scale-2006

Table 3.1 continued

4	Test anxiety	Multidimensional scale	Putwain et al. (2020)	The Development and Validation of a new Multidimensional Test Anxiety Scale(MTAS)
5	Teachers' Efficacy	TSES scale	Dupuis et al. (2020)	Relationship between Self-Efficacy Measured by the TSES Scale and Teacher Participation in PDS Activity
6	Peer relationship	MSLSS scale	Sawatzky et al. (2009)	Sample heterogeneity and the measurement structure of the Multidimensional Students' Life Satisfaction Scale. Social Indicators Research, 94, 273-296

The use of a survey instrument facilitates the acquisition of data that has the potential to provide valuable insights pertaining to the research inquiry (Pumptow & Brahm, 2020). Within the framework of this study, which seeks to comprehend the individual, family, school, and social dimensions that impact student performance (Xie & Zhang, 2020; Zafari et al., 2021), questionnaires were used as the principal means of data collection of some factors.

Initially, the existing current data in high school F is as follows (see Table 3.2).

Table 3.2: The Existing Data in High School F

Section	School's existing database
Name, Age, and gender	Included in the past performance information sheet
Past performance A: admission performance in 2022	No.1. Admission performance in 2022 (including subjects' performance and total scores in 2022). ("Total scores" refers to the overall score of all subjects in one semester)
Past performance B: high school performance, and NCEE performance	No.2. Mid-term Examination performance before subject options in 2022 (including subjects' performance and total scores)
	No.3. Final-term Examination performance before subject options in 2023 (including subjects' performance and total scores)
	No.4. Mid-term Examination performance after subject options in 2024 (including subjects' performance and total scores)
	No.5. NCEE performance in 2025 (including subjects' performance and total scores)

Note: The number appearing before the subject in the later part indicates the number of the examination.

The assessment of individual and family factors influencing students' academic performance may be conducted by surveys administered to students themselves. Similarly, the evaluation of school and social elements impacting students' academic performance can be gained through surveys administered to students and teachers. The resulting data is shown in the tables provided below (see Table 3.3 - 3.4). The scores are directly quoted from the survey questionnaires in several original references, and the classification into Low, Moderate, and High levels is based on specific score ratios.

Table 3.3: Data on Individual and Family Factors Affecting Student Performance

Section	Questionnaire for Students	Sub-Section		
Individual Factors	Age (Existing data)	Based on the age of entry in 2022		
	Gender (Existing data)	Male and female		
	Motivation level	Level	Score	Interpretation
		Low	1-36	Disinterested, passively participates
		Moderate	37-72	Follows along, occasional distraction
		High	73-108	Actively engaged, enthusiastic
	Self-efficacy	Level	Score	Interpretation
		Low	1-67	Doubts abilities, avoids challenges
		Moderate	68-133	Confident in some tasks, seeks help when needed
		High	134-200	Believes in capabilities, embraces challenges
	Test anxiety	Level	Score	Interpretation
		Low	1-27	Calm during tests, performs well
		Moderate	28-53	Nerves before exams, but manages to focus
		High	54-80	Panics during tests, struggles to concentrate
	Past performance (Existing data)	Level	Score	Interpretation
		Low	<0.6	Below 60% of aggregate score
		Moderate	$\geqslant 0.6$ and < 0.8	More than 60% of aggregate score (including 60%) and less than 80%
		High	$\geqslant 0.8$	More than 80% of aggregate score (including 80%)
Family Factors	Father's education level	Level	Interpretation	
		Low	Junior high school and below	
			Senior high school	
		Moderate	Undergraduate degree	
		High	Master's degree	
			Doctoral degree	

Table 3.3 continued

	Level	Interpretation	
Mother's education level	Low	Junior high school and below	
		Senior high school	
	Moderate	Undergraduate degree	
	High	Master's degree	
		Doctoral degree	
Family income (annual)	Level	Interpretation	
	Low	<10000	
	Moderate	10000-30000	
		30000-60000	
		60000-100000	
	High	>100000	
Parental involvement in student's learning	Level	Score	Interpretation
	Low	1-16	Minimal interaction, rarely checks progress
	Moderate	17-32	Monitors progress periodically, provides moderate guidance
	High	33-48	Actively engaged, helps with homework and others regularly

Table 3.4: Data on School and Social Factors Affecting Student Performance

Section	Questionnaire for Students and Teachers	Sub-Section		
School Factors	Teacher's education level	Level	Interpretation	
		Low	Undergraduate degree	
		Moderate	Master's degree	
		High	Doctoral degree	
	Teacher's qualification	Level	Interpretation	
		Low	Secondary school teacher II	
		Moderate	Secondary school teacher I	
		High	Secondary school associate senior teacher	
			Secondary school full senior teacher	
	Teaching methods		Interpretation	
			Traditional teaching method	
			Smart classroom teaching methods	
			Both	
	Teacher's self-efficacy	Level	Score	Interpretation

Table 3.4 continued

		Low	1-20	Teachers lack confidence, feel stressed, and avoid challenges
		Moderate	21-40	Teachers have moderate confidence, manage tasks, but feel uneasy with complexity
		High	41-60	Teachers are confident, adaptable, and inspire student engagement
Social Factors	Peer relationships	Level	Score	Interpretation
		Low	1-9	Doubts abilities, lacks confidence in teaching
		Moderate	10-19	Cautiously confident, engages supportively when necessary
		High	20-28	Strongly believes in effectiveness, highly motivated
	Social support	Yes		
		No		
	Private tutoring	Yes		
		No		
	Socioeconomic status (Best on family income and parental education level)	Level	Score	Interpretation (Family income, father's education level, mother's education level: high = 3, moderate = 2, low = 1)
		Low	3-4	Low income and parents with limited education
		Moderate	5-6	Family income and parental education level are generally moderate
		High	7-9	High income and parents with higher education degrees

3.3.3 Data Cleaning

Within the domain of scholarly inquiry, data assumes a pivotal role, serving as the fundamental basis upon which insights, findings, and recommendations are derived (Muthalagu et al., 2014). Nevertheless, raw data, particularly when obtained from many sources, sometimes exhibits inherent flaws. If these defects are not addressed, they have the potential to introduce bias into the findings, hence resulting in erroneous or even deceptive conclusions (Kim et al., 2019). Schultze et al. (2021) also highlight the utmost significance of the data cleansing procedure.

a) Understanding Data Cleaning

The nature of defects in data may emerge in a multitude of ways. The types of data quality issues encompass missing values, inconsistencies, redundancies, and anomalies (Bao et al., 2022). Missing values refer to the absence of certain data points. Inconsistencies arise when data contradicts other data points or expected patterns. Redundancies occur when the same data is repeated. Anomalies denote data points that significantly deviate from anticipated patterns or norms.

Rectification strategies refer to the methods or approaches used to correct errors, mistakes, or inaccuracies in various contexts. Once problems are discovered, measures for repair are implemented. To address missing values, one common approach is imputation, which is replacing missing data points with approximated values. To address any discrepancies, it may be necessary to conduct a more comprehensive analysis of the data source to identify the underlying factors and rectify them accordingly. The mitigation of redundancies may be achieved by using deduplication techniques, while the identification of abnormalities may need the use of outlier detection methods (Lv et al., 2023).

b) Data Cleaning in the Context of the Study

The study primarily focuses on NCEE and its associated repercussions, thereby necessitating a wide range of data sources. This encompasses the following:

The use of online questionnaires has become more prevalent in academic study and data collection. Although digital technologies are often efficient, they may sometimes result in partial or inconsistent replies. For example, a student may choose to omit certain questions or provide responses that are inconsistent with one another. It is important to address these concerns to protect the integrity of the data.

Academic databases are comprehensive collections of scholarly resources that provide access to a wide range of academic materials, including journal articles, research papers. The academic database of high school F, albeit a valuable repository of knowledge, presents several issues that are unique to its nature. Historical data may exhibit instances of missing entries or inconsistencies in the recording of data over different time periods.

The data cleaning technique for this study is notably complex due to the diverse range of sources used. The task at hand involves not only guaranteeing the precision of individual data points but also ensuring that the data, when examined comprehensively, presents a coherent and dependable depiction of whole landscape (Ngito & Kwaba, 2022).

Data cleaning is not only a procedural component of the study process. The current phase is of utmost importance as it serves to ascertain the legitimacy, validity, and dependability of the study. Through a thorough examination of data inadequacies, this study establishes a foundation for generating insights that are resilient and accurately represent the essence of NCEE and its ramifications.

c) Feature Selection

The process of feature selection plays a pivotal role in the data analysis pipeline, particularly when the objective is to extract significant insights or make accurate predictions based on the data. The significance of this stage in predicting student performance (Tsang et al., 2020; Yan & Liu, 2020), particularly in high-stakes assessments such as the NCEE, cannot be overemphasised.

The purpose of this section is to get a comprehensive understanding of feature selection techniques. The primary objective of feature selection is to discern and preserve just those characteristics or indications that possess a substantial influence on the desired result, namely, the academic performance of students. By using this approach, it guarantees that the constructed models possess both accuracy and interpretability, as well as efficiency (Michel et al., 2021).

The acceleration of model training and reduction in computing resource requirements may be achieved by the process of dimensionality reduction in the dataset, as shown by Salam et al. (2021). The inclusion of irrelevant or duplicate characteristics in the data might contribute extraneous information, which may result in overfitting. According to Nasution et al. (2018), the enhancement of the model's accuracy may be achieved by removing these factors. The interpretability of models may be enhanced by constructing them using a limited number of relevant attributes. This characteristic facilitates comprehension and enables stakeholders to adopt more effective actions (Hong et al., 2020).

Filter methods are a class of techniques that assess the importance of characteristics using statistical measurements and thereafter keep only those aspects that are ranked highest. Several commonly used procedures in statistical analysis include correlation coefficients, chi-square tests, and information gain (Santos-Rufo et al., 2020). Wrapper methods are a class of feature selection techniques that include evaluating subsets of features by training a model on them and measuring its performance. The category in question includes techniques such as backward elimination and forward selection (Li & Wang, 2015). Embedded methods are a kind of feature selection technique that incorporates the process of selecting relevant features into the model training process.

The study is centred on the use of EDM methods to predict student performance in NCEE. The data utilised in this research encompasses a wide array of sources, including online surveys and academic databases. The presence of variety within a given context suggests the existence of a wide range of possible characteristics, each of which provides a distinct viewpoint on the academic performance of students (Hamoud & Humadi, 2019).

Demographic data, such as age, gender, and socioeconomic background, may provide valuable insights on the academic performance of students (Vargas-Ramos et al., 2021). The academic history of an individual, including their past performance, subject options, and involvement in extracurricular activities, has been suggested as a potential indicator of their future success (Sahni, 2023). The feedback obtained via questionnaires may provide valuable insights into several aspects, including areas of concern, techniques for preparation, and perceived problems. This information is derived from responses received from both students and teachers (Tuffah & Al-Jubouri, 2021).

The significance of the feature selection procedure becomes paramount due to the abundance of possible characteristics (Bertolini et al., 2021). The inclusion of this step guarantees that the succeeding phase of model construction is based on data that is highly representative of student performance, hence maximising the likelihood of obtaining significant and practical insights (Disha & Waheed, 2022). The process of feature selection has significance beyond its technological implementation inside the data analysis procedure. The choice made regarding strategy has a significant role in determining the trajectory and results of the study (Taguchi & Turki, 2021). Through a laborious process of carefully identifying the most relevant and influential variables, this study aims to provide a comprehensive and nuanced comprehension of the factors that affect student performance in NCEE.

3.3.4 Model Building

The construction of a model plays a crucial role in the study process, particularly when the aim is to extract predicted insights from collected data. Within the realm of educational research, where the consequences are significant and the ramifications extend far, the selection of modelling methodologies and their execution has heightened importance (Merema et al., 2019).

The data obtained from Excel was afterwards transferred to the EDM tool, namely Orange, to do additional analysis in this study. Prior to producing predictions, a thorough examination was conducted on all factors that have an impact on the students to ascertain their effect on the precision of the deposition predicts in the subjects of Chinese, Mathematics, and English. In the first stages, a comprehensive analysis was conducted to include all relevant factors in formulating the prediction. Subsequently, the attribute exhibiting the least weight was deleted to assess its impact on the accuracy of predictions. If there was an increase in prediction accuracy after the removal of the characteristic, it was then excluded from any future study. The technique was iteratively performed until the highest level of prediction accuracy was achieved. Subsequently, the model will be generated. This section demonstrates the use of EDM technology for the purpose of predicting student performance. The model was constructed using an 8:2 ratio for training and testing, implying that 80% of the data was allocated for training purposes, while the remaining 20% was reserved for testing （Alboaneen et al, 2022）. The training dataset was used to train the EDM algorithm, whilst the test dataset was employed to evaluate the performance of the EDM approach. The use of the test set was employed to evaluate the performance of the classifier on data that had not been encountered earlier. The figure 3.2 shown below illustrates the procedural steps involved in the construction of an EDM model, as described by Roslan and Chen (2023).

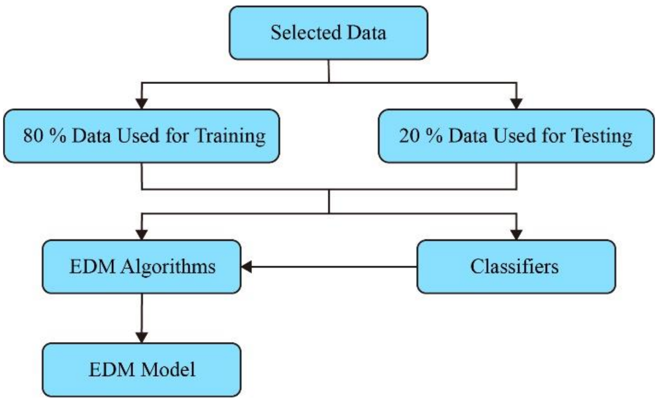

Figure 3.2: The Process of Constructing an EDM Model

a) The Essence of Model Building

The main objective of constructing a model is to provide a depiction of the fundamental patterns and connections contained in the data. The objective of this study is to predict students' academic performance and recommend their subject options. The selection of algorithms is determined by several factors, including the characteristics of the data, the research goals, and the intended results. Several often-used strategies in educational research include Naïve Bayes (NB), Decision Tree (DT), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs) (refer to Table 3.5), according to the study conducted by Alsariera et al. (2022).

Table 3.5: Mechanical Learning Algorithms Commonly Used within Education

Section	Learning Task	Common Algorithms
Supervised Learning	Classification	Naïve Bayes Decision Tree Artificial Neural Networks Support Vector Machines

①Naïve Bayes (NB) method is a probabilistic classification strategy that applies Bayes' theorem, assuming independence among predictors. In basic words, a NB classifier operates on the assumption that the occurrence of a certain characteristic within a class is independent of the occurrence of any other feature. Below is a concise explanation of the algorithm. Bayes' Theorem is a fundamental concept in probability theory and statistics. The NB method relies on Bayes' theorem, a mathematical theory used to calculate probabilities by incorporating prior probabilities, and this theorem offers a mathematical foundation for updating the probability of a basic component of the NB algorithm. (Ulfah, F., 2022). The formula is as follows.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Equation 3.1

Where:

• P(A｜B) is the probability of hypothesis A given the data B.

• P(B｜A) is the probability of data B given that the hypothesis A was true.

• P(A) is the probability of hypothesis A being true (regardless of the data). This is known as the prior probability of A.

• P(B) is the probability of the data (regardless of the hypothesis).

The approach first calculates the prior probability for each class included in the dataset. Afterwards, the system calculates the probability of features for each class, so guaranteeing a thorough comprehension of feature probabilities. By using Bayes' Theorem, the system proceeds to compute the posterior probability for each class. In order to provide a coherent and rational decision-making process, the prediction is made by selecting the class with the greatest posterior probability. NB classification model is as follows.

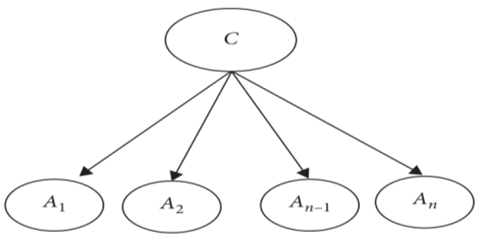

Figure 3.3: NB Classification Model (Xia & Yan, 2021)

②DT is a prevalent kind of supervised machine learning algorithm mostly used for classification tasks (Charbuty & Abdulazeez, 2021). DT is graphical representations that has resemblance to flowcharts. Within these structures, the internal nodes are used to represent characteristics, the branches are used to symbolise decision rules, and the leaf nodes are employed to represent outcomes. The commencement of the expedition takes place at the root node. The dataset is segmented by the algorithm according to attribute values. The segmentation procedure is conducted iteratively, resulting in the formation of a hierarchical tree structure that consists of linked nodes and branches. The process is executed in the following manner. The identification of the most appropriate attribute for data partitioning is accomplished via the use of attribute selection measures. This characteristic then divides the dataset into many subgroups. Every subgroup follows an identical procedure until certain predetermined conditions are satisfied. The determination of the characteristic for each node is thereafter conducted. During the final stage, tree branches that demonstrate inadequate classification performance are trimmed to improve the efficiency and clarity of the model. DT model is as follows.

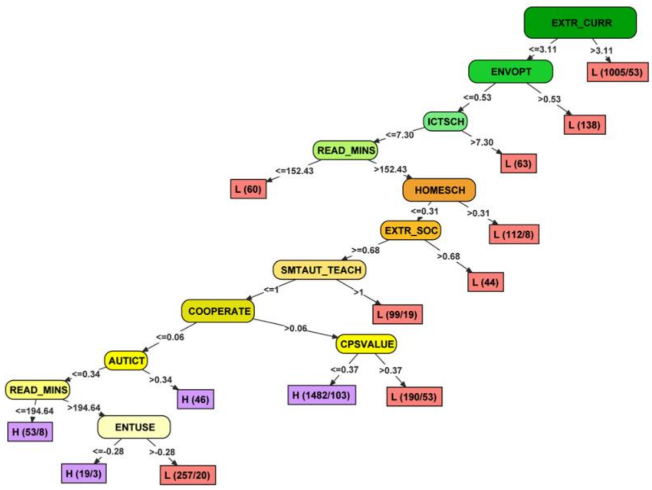

Figure 3.4: DT Model (Martínez-Abad et al, 2020)

③Artificial Neural Networks (ANNs) are computer models that draw inspiration from the structural organisation of the human brain (Kaveh & Mesgari, 2022). ANNs demonstrate a high level of proficiency in pattern recognition and predictive capabilities, especially when they undergo training using large and comprehensive datasets. The operational flow of ANNs may be described as follows. At the outset, the weights and biases of all neurons are initialised with random values. The computational process of the network starts at the input layer, where data is received and processed. The data thereafter passes via one or more concealed layers, where it receives further processing using activation functions. The data that has undergone processing reaches its last stage at the output layer, leading to the generation of a prediction or categorization. If disparities emerge between the anticipated and observed outputs, the backpropagation technique intervenes to calculate the error and optimise the weights and biases. The iterative procedure including forward, and backward passes continues until the error reaches a satisfactory threshold or a preset number of iterations is attained, so guaranteeing the development of a more polished and coherent model. The adaptive prediction paradigm for online education, exemplified by ANNs, may be described as follows.

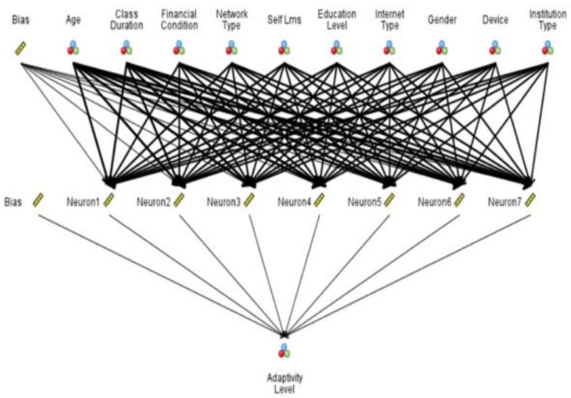

Figure 3.5: ANNs Model (Malik et al, 2023)

④Support Vector Machines (SVMs) are supervised machine learning techniques that are capable of addressing classification and regression tasks. The classification process involves the identification of a hyperplane that optimally separates a given dataset into distinct classes (Gao et al., 2021). The algorithm process refers to the systematic set of steps or instructions that are followed to solve a problem or achieve a certain goal. SVMs are widely recognised as a prominent supervised learning techniques that demonstrate proficiency in performing classification and regression tasks. The core principle of SVMs is in its capacity to identify the most ideal hyperplane that effectively separates data points into various classes. This characteristic makes SVMs very advantageous in the realm of student performance prediction. This extensive analysis explores the fundamental principles behind SVMs. The fundamental purpose of a kernel function is to facilitate the transformation of data into a higher-dimensional space. In this context, SVMs diligently ascertain the ideal hyperplane that effectively separate classes by maximising the margin between them. Support vectors, which are data points located in close vicinity to the hyperplane, serve a crucial role in identifying its location. The categorization of new data points is determined by their relative location to the hyperplane when further data points are introduced. Significantly, in situations when the data does not possess intrinsic linear separability, the kernel function effectively performs a transformation on the data, hence enhancing the resilience and adaptability of the SVMs algorithm. A kind of SVMs model is as follows.

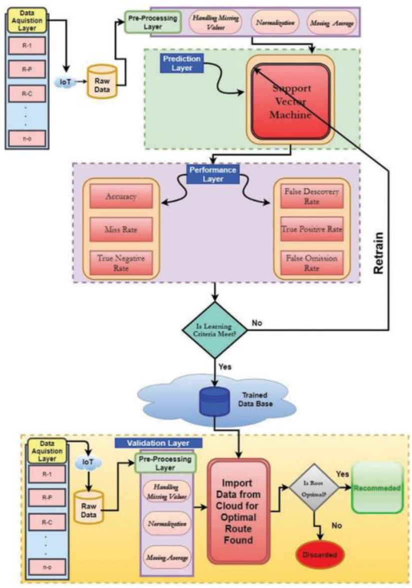

Figure 3.6: SVMs-CV2X-M4 System Model (Khan et al, 2022)

b) Model Building in the Context of the Study

The study’s primary emphasis on NCEE and its associated ramifications renders the data abundant and diverse. This encompasses several types of data, such as academic performance measures, responses obtained through surveys, demographic characteristics, and other relevant information. Hence, it is essential for the model construction procedure to adequately consider its intricate nature.

Feature integration occurs after the feature selection process, wherein the selected characteristics are included into the model. This practise guarantees that the model has access to all pertinent information required for generating precise predicts (Behnke et al., 2021). The process of training and validation is an essential component in the field of machine learning. After establishing the model, it undergoes training using a portion of the available data. The training procedure encompasses the provision of data to the model, enabling it to adapt its internal parameters to optimise its alignment with the provided data. Following the completion of the training process, the model undergoes validation using a distinct dataset to evaluate its precision and dependability (Alsharif & Rawat, 2021).

The phase of model construction necessitates a rigorous approach, demanding a comprehensive comprehension of both the data and the techniques used. The process involves the conversion of unprocessed data into practical and valuable insights, equipping individuals with the necessary resources to make well-informed options (Moon & Alabdulwahab, 2020). Within the framework of NCEE, the use of a meticulously constructed model may provide immeasurable value by furnishing educators and students with important perspectives, enabling them to effectively traverse the intricacies inherent in the examination process. The process flowchart for the framework is shown as follows.

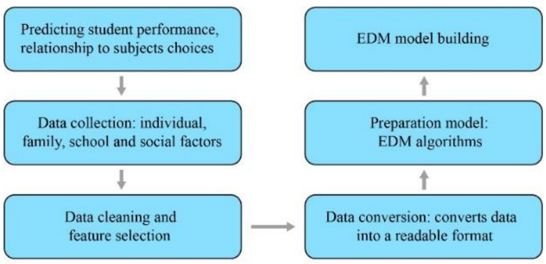

Figure 3.7: The framework Process Flowchart

3.3.5 Result Evaluation

Result evaluation is an essential stage in the study process, particularly in the context of predictive modelling (Mi et al., 2022). The statement highlights the need of a checkpoint in the process of model building. This checkpoint not only assures the mathematical accuracy of the models but also emphasises their practical relevance and ability to be implemented effectively (Pandey & Sharma, 2013). The evaluation process has significant importance in the context of predicting student performance and test readiness, since it is characterised by high stakes (Hu, Ma, & Chen, 2021).

According to Yen et al. (2014), this function acts as a critical evaluation point to verify that the constructed models are not only theoretically valid but also possess practical relevance and can be effectively implemented. Within the framework of predicting student performance and assessing test readiness, the stakes are considerably elevated, hence amplifying the significance of the assessment procedure.

Considering the study's specific emphasis on NCEE and its associated ramifications, it is essential to provide a thorough review procedure. The constructed models have the objective of predicting significant outcomes such as student performance and subject options. The presence of mistakes or discrepancies in projections may have substantial ramifications for students, educators, and institutions. The implications for stakeholders are significant, as accurate predicts have the potential to inform educators in customising interventions, delivering focused assistance to students, and enhancing instructional techniques (Nguyen-Huy et al., 2022). Accurate predicts may provide students with valuable information into their areas of proficiency and areas that need improvement, so assisting them in formulating effective preparation methods (Gong et al., 2015).

Result evaluation is a continuous and ongoing process, rather than a one-time occurrence. Based on the results of the assessment, it is possible that the models may need further refining. The iterative nature of this procedure guarantees the ongoing relevance and accuracy of the models considering newly available data (Staniszewska et al., 2021). Result evaluation is not just a technical component inside the research process. The process serves to guarantee the legitimacy, relevance, and actionability of the research. Through a thorough and rigorous examination of the outcomes, this study guarantees that its conclusions and suggestions possess both strong validity and relevance to the practical complexities and intricacies of NCEE environment.

3.3.6 Deployment

The deployment phase is the last stage in which the outcomes of thorough study, data analysis, and model development are incorporated into practical applications within the actual world. Within the framework of this research, the term "deployment" refers to the integration of predictive models into the educational environment of high school F, with the objective of improving the subject options of students.

The concept of continuous monitoring and feedback loop refers to the ongoing process of systematically observing and evaluating a particular system or process and providing regular feedback to improve its performance. The act of deployment is not a singular occurrence. The process is characterised by its ongoing nature, necessitating the consistent monitoring, provision of feedback, and iterative adjustments. As the models undergo evaluation in practical contexts, various insights, obstacles, and prospects are likely to arise. According to Schenke et al. (2017), the implementation of a feedback loop facilitates the integration of these insights back into the system, hence fostering a state of ongoing improvement.

Considering the primary objective of this study, which is to predict student performance and subject options for NCEE, the deployment phase has significant significance. The ideas and insights have the capacity to fundamentally transform the educational approach of high school F. Deployment is the critical phase when practical implementation and execution take place. This phase represents the activation of complex models and algorithms, which have a profound influence on students' lives, moulding their academic paths and propelling them towards performance. The synergy of data scientists, educators, administrators, and students guarantees that the implementation is not only technically robust but also pedagogically relevant and influential.

In summary, the research design draws inspiration from EDM approach proposed by Malini and Kalpana (2021). This design provides a structured and all-encompassing technique for comprehending and predicting student performance within the framework of NCEE. The primary objective of this section is to provide practical and valuable insights that may effectively improve the academic experiences and results of students in China, using a systematic approach consisting of clearly defined phases.

3.4 Data Collection Procedures

In the realm of educational research, the procedures for data collection are paramount to ensure the integrity, accuracy, and relevance of the gathered information. For this study, which sought to understand the factors influencing student performance in the context of NCEE, a meticulous data collection procedure was employed.

Given the unique position of the study case school, which enjoyed a significant degree of decision-making autonomy, the initial step involved securing the necessary permissions. Recognizing the importance of adhering to ethical standards and ensuring the confidentiality of the data, the researcher first gained approval from their academic supervisor. This was a crucial step, ensuring that the study was in line with academic standards and ethical considerations. Following this, the researcher drafted an application letter detailing the objectives, methodologies, and potential implications of the study. This letter was addressed to the case school, providing them with a comprehensive overview of the study. To further bolster the application, a letter of authorization from the faculty was attached, underscoring the academic legitimacy of the research endeavour (see Appendix C). Upon receiving the application, the case school's headmaster reviewed the proposal, weighing the potential benefits against any potential risks. Given the non-intrusive nature of the study and the potential insights it promised, the headmaster granted permission for the data collection. With the necessary permissions in place, the researcher proceeded to gather data. This involved accessing the school's academic databases, administering online questionnaires to students and teachers, and conducting observational studies. The data encompassed a wide range of variables, from individual student characteristics to broader school and social factors. Once the data was amassed, it was prepared for analysis. Given the study's emphasis on leveraging EDM techniques, the raw data underwent rigorous cleaning and preprocessing. This ensured that the data was consistent, accurate, and ready for advanced analytical procedures.

The data collection procedures (see Figure 3.8) for this study were characterised by meticulous planning, adherence to ethical standards, and a collaborative approach with the case school. The result was a rich dataset, primed for analysis using EDM techniques, promising valuable insights into the factors influencing student performance in the context of NCEE.

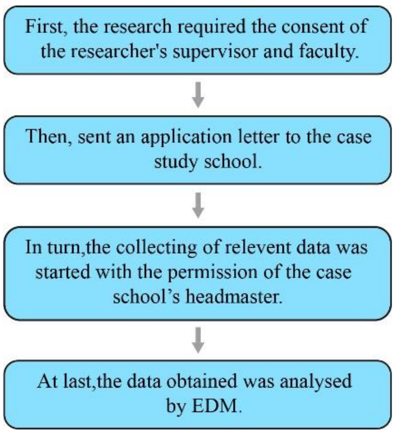

Figure 3.8: Data Collection Procedures

3.5 Research Ethics

The adherence to research ethics is a fundamental aspect of any academic pursuit, as it guarantees the execution of study with principles of honesty, consideration, and accountability (McGuire, 2008). Within the framework of this study endeavour, which aims to comprehensively comprehend the determinants impacting academic performance among students in the setting of NCEE, significant attention was dedicated to two fundamental ethical considerations: informed consent, and the safeguarding of privacy and data integrity.

The concept of informed permission refers to the ethical practise of obtaining consent from someone after providing them with all relevant information about a certain activity. The notion of informed consent highlights the significance of obtaining express agreement from all those engaged in the study (Dunlap, 1998). Prior to commencing the study endeavour, a comprehensive strategy was implemented to get the necessary licences. The researcher first sought and gained consent from their academic supervisor to ensure that the research was in line with the larger academic objectives and ethical standards of the institution. Following that, the academic staff issued a formal letter of authorisation, so enhancing the legitimacy of the study undertaking.

Having obtained these permissions, the researcher proceeded to engage with the headmaster of the case school (see Appendix D). Considering the considerable level of decision-making autonomy possessed by the educational institution, it was crucial to provide a thorough exposition of the study aims, techniques, and prospective ramifications (Berkman et al., 2018). The endorsement of the headmaster served not only as a mere administrative need, but also as a testimonial to the collective cooperation and synergy shown in the study endeavour. In addition, the participants of the study, namely the students and instructors, were also contacted and their informed permission was gained. This measure was used to guarantee that all participants had comprehensive knowledge on the extent and ramifications of the study and made voluntary decisions to participate in the study.

The protection of data and the preservation of persons' privacy have become prominent concerns in the current era of digital technology (Liu et al., 2015). Considering this recognition, the researcher implemented rigorous protocols to safeguard the confidentiality and integrity of the obtained data. The complete collected data is stored securely on a password-protected computer hard drive, where it will be retained indefinitely for ongoing access and use. The identity of the research participants was not divulged at any stage, and any sensitive material was appropriately anonymised. The use of this measure not only served to safeguard the participants from possible risks but also upheld the credibility and reliability of the study outcomes (Zhang et al., 2020).

In addition to safeguarding the privacy of the individuals involved, ensuring the security of the data was of utmost significance. The proliferation of technology has led to an amplification of the potential dangers associated with unauthorised access and data breaches (Mustafa et al., 2022). To mitigate these potential risks, the researcher implemented stringent data security protocols. This included the use of encrypted storage systems, safe data transport methods, and periodic security assessments. These procedures were implemented to safeguard the protection of data against unauthorised access and possible abuse.

The ethical issues of this study were not perfunctory tasks, but rather fundamental concepts that informed and guided every phase of the investigation. By placing a strong emphasis on obtaining informed consent and ensuring data protection, the study not only upheld the utmost ethical principles but also cultivated a climate of trust and cooperation among all parties involved.

3.6 Chapter Summary

The present chapter provides an overview of the methods used in the conducted study. The text provides an overview of the research setting, research methodology, including the several key phases of EDM, the survey tool used, and considerations pertaining to research ethics. The next chapter includes the presentation of data analysis and results.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.1 Introduction

This chapter presents the results derived from data analysis. The data analysis process encompasses data preparation, data preprocessing, and data transformation, followed by model building and evaluation, to address the research questions.

4.2 Data Source Results

After obtaining approval from the supervisor, the faculty, and the high school principal, the researcher collected the data. The scope of the data was broad. Some data came from the existing database of the high school. The rest was gathered through questionnaires completed by students and teachers who agreed to participate in the survey.

The existing school data included individual factors such as student name, age, gender, and past examinations performance. The remaining data was collected through questionnaires. These questionnaires covered individual factors like motivation level, self-efficacy, and test anxiety. They also included family factors such as father's education level, mother's education level, annual family income, and parental involvement in student's learning. School factors included teacher's qualification, teacher's education level, teaching methods, and teacher's self-efficacy. Social factors involved peer relationship, social support, private tutoring, and socioeconomic status.

A total of 1,127 student records and 88 teacher records were collected after data cleaning. In summary, the collected data covered four dimensions: individual, family, school, and social factors.

4.2.1 Individual Dimensions

a) Age and gender

Table 4.1 lists the numbers and proportions of students of various age and gender, based on the 2022 student enrolment information form. The student age is based on his or her age at the time of enrolment in 2022.

Table 4.1: The Numbers and Proportions of Students of Various Age and Gender

Items	Attributes	Numbers	Proportions (%)
Age	14 years old	41	3.64
	15 years old	786	69.74
	16 years old	287	25.47
	17 years old	10	0.89
	18 years old	1	0.09
	20 years old	2	0.18
	Total	1127	100
Gender	Female	537	47.65
	Male	590	52.35
	Total	1127	100

Table 4.1 shows that most students entered school at the age of 15 or 16. Among them, 69.74% were 15 years old, and 25.47% were 16 years old. Students of other ages made up a small percentage. Regarding gender, 47.65% were female, and 52.35% were male, with slightly more male students than female students.

b) Motivation level, self-efficacy and test anxiety

Table 4.2 shows the results of students' motivation level, self-efficacy, and test anxiety from the questionnaire.

Table 4.2: Motivation Level, Self-efficacy and Test Anxiety Results

Items	The number of students (Percentage %)		
	High	Moderate	Low
Motivation level	280 (24.84)	228 (20.23)	619 (54.92)
Self-efficacy	648 (57.49)	228 (20.23)	231 (20.49)
Test anxiety	643 (57.05)	229 (20.32)	255 (22.63)

From Table 4.2, about half of the students have a moderate motivation level, accounting for 54.92%. This is followed by 28.84% with a high motivation level and 20.23% with a moderate motivation level. For self-efficacy, more than half of the students have high self-efficacy, about 57.49%. Next, 20.23% have moderate self-efficacy, and about 20.49% have low self-efficacy. For test anxiety, more than half of the students have high test anxiety, about 57.05%. Around 20.32% have moderate test anxiety, and a similar percent, 22.63%, have low test anxiety.

c) Past performance

For the factor of past performance, Tables 4.3 to 4.7 summarize the scores of each subject and total scores in five major examinations: No.1 - student admission performance (only six subjects were covered), No.2 - examination performance before subject options in 2022, No.3 - examination performance before subject options in 2023, No.4 - examination performance after subject options in 2024, and No.5 - NCEE performance in 2025.

Table 4.3 details students' academic proficiency upon admission, categorized into high, moderate, and low performance across various subjects. In Chinese, 224 students (19.88%) achieved high performance, 229 (20.32%) demonstrated moderate performance, and a majority of 674 (59.80%) fell into the low category. For Mathematics, 252 students (22.36%) scored high, while 209

(18.54%) and 666 (59.09%) were in the moderate and low ranges, respectively. In English, 243 students (21.56%) performed at a high level, 229 (20.32%) at a moderate level, and 655 (58.12%) at a low level. Physics results indicated 232 high performance (20.59%), 201 moderate (17.83%), and 694 low (61.58%). Chemistry had 209 students (18.54%) in the high group, 260 (23.07%) in the moderate, and 658 (58.39%) in the low group. In Politics, 245 students (21.74%) performed highly, 209 (18.54%) moderately, and 673 (59.72%) poorly. History showed a slightly different trend, with 349 students (30.97%) in the high category, 128 (11.36%) moderate, and 650 (57.68%) low. Finally, analysis of total points showed that 242 students (21.47%) achieved high total scores, 217 (19.25%) were moderate, and 668 (59.27%) were low. Overall, the proportion of low-performing students was consistently around 58%–61% in most subjects, indicating a widespread performance issue at admission.

Table 4.3: Student Admission Performance

Subjects	**The number of students (Percentage %)**		
	High	**Moderate**	**Low**
Chinese	224 (19.88)	229 (20.32)	674 (59.80)
Mathematics	252 (22.36)	209 (18.54)	666 (59.09)
English	243 (21.56)	229 (20.32)	655 (58.12)
Physics	232 (20.59)	201 (17.83)	694 (61.58)
Chemistry	209 (18.54)	260 (23.07)	658 (58.39)
Politics	245 (21.74)	209 (18.54)	673 (59.72)
History	349 (30.97)	128 (11.36)	650 (57.68)
Total scores	242 (21.47)	217 (19.25)	668 (59.27)

Table 4.4 illustrates performance shifts before subject options in 2022, reflecting increasing academic demands. In Chinese, 229 students (20.32%) scored high, 250 (22.18%) scored moderately, and 648 (57.50%) scored low. For Mathematics, 237 students (21.02%) were in the high category, 227 (20.14%) in the moderate, and 663 (58.83%) in the low. In English, 235 students (20.85%)

achieved high marks, 224 (19.88%) moderate, and 668 (59.27%) low. Physics scores showed that 234 students (20.76%) performed highly, 229 (20.32%) moderately, and 664 (58.92%) poorly. Chemistry results revealed 237 high scorers (21.03%), 223 moderate (19.79%), and 667 low (59.18%). For Biology, the high-performance group included 253 students (22.45%), the moderate 234 (20.76%), and the low 640 (56.79%). In Politics, 224 students (19.88%) achieved high scores, 267 (23.69%) were moderate, and 636 (56.43%) low. History presented 274 high scorers (24.31%), 211 moderate (18.72%), and 642 low (56.97%). Geography had the highest proportion of high scorers: 285 (25.29%), with 221 moderate (19.61%) and 621 low (55.10%). Regarding total scores, 228 students (20.23%) were in the high category, 225 (19.96%) in the moderate, and 674 (59.80%) in the low.

Table 4.4: Examination Performance before Subject Options in 2022

Subjects	**The number of students (Percentage %)**		
	High	**Moderate**	**Low**
Chinese	229 (20.32)	250 (22.18)	648 (57.50)
Mathematics	237 (21.02)	227 (20.14)	663 (58.83)
English	235 (20.85)	224 (19.88)	668 (59.27)
Physics	234 (20.76)	229 (20.32)	664 (58.92)
Chemistry	237 (21.03)	223 (19.79)	667 (59.18)
Biology	253 (22.45)	234 (20.76)	640 (56.79)
Politics	224 (19.88)	267 (23.69)	636 (56.43)
History	274 (24.31)	211 (18.72)	642 (56.97)
Geography	285 (25.29)	221 (19.61)	621 (55.10)
Total scores	228 (20.23)	225 (19.96)	674 (59.80)

The 2023 examination results before subject options reveal distinct challenges (see Table 4.5). In Chinese, 274 students (24.31%) attained high scores, 226 (20.05%) scored at a moderate level, while 627 (55.63%) fell into the low-performance group. Mathematics outcomes indicated 239 students (21.21%) with high achievement, 220 (19.52%) moderate, and 668 (59.27%) low. English scores were lower, with only 223 students (19.79%) categorized as high, 231 (20.50%) moderate, and 673 (59.72%) low. For Physics, 231 students (20.50%) achieved high performance; 218 (19.34%) moderate; and 678 (60.16%) low. Chemistry results revealed 246 students (21.83%) in the high range, 211 (18.72%) moderate, and 670 (59.45%) low. Biology showed slightly better results: 254 students (22.54%) performed well, 220 (19.52%) were moderate, and 653 (57.94%) were low achievers. Politics had the highest proportion of high performers, with 287 students (25.47%) in that group, followed by 229 (20.32%) moderate and 611 (54.21%) low. In History, 246 students (21.83%) reached high performance, 207 (18.37%) moderate, and 674 (59.80%) low. Geography results showed 257 students (22.80%) high, 222 (19.70%) moderate, and 648 (57.50%) low. Finally, total scores revealed 224 students (19.88%) high, 225 (19.96%) moderate, and 678 (60.16%) low.

Table 4.5: Examination Performance before Subject Options in 2023

Subjects	**The number of students (Percentage %)**		
	High	**Moderate**	**Low**
Chinese	274 (24.31)	226 (20.05)	627 (55.63)
Mathematics	239 (21.21)	220 (19.52)	668 (59.27)
English	223 (19.79)	231 (20.50)	673 (59.72)
Physics	231 (20.50)	218 (19.34)	678 (60.16)
Chemistry	246 (21.83)	211 (18.72)	670 (59.45)
Biology	254 (22.54)	220 (19.52)	653 (57.94)
Politics	287 (25.47)	229 (20.32)	611 (54.21)
History	246 (21.83)	207 (18.37)	674 (59.80)
Geography	257 (22.80)	222 (19.70)	648 (57.50)
Total scores	224 (19.88)	225 (19.96)	678 (60.16)

Following the subject options in 2024, student performance remained uneven across disciplines (see Table 4.6). In Chinese, 256 students (22.72%) reached the high-performance tier, while 211 (18.72%) were placed in the moderate group, and 660 (58.56%) fell into the low category. Mathematics results showed 231 high achievers (20.50%), 220 moderate (19.52%), and 676 low performance (59.98%). For English, 225 students (19.96%) attained high performance, with 217 (19.25%) performing moderately and 685 (60.78%) recording low performance. Physics had 152 high-performing students (20.13%), 149 moderate (19.74%), and 454 low (60.13%). In Chemistry, the high group comprised 152 students (20.13%), the moderate 153 (20.26%), and the low 450 (59.60%). Biology results reflected a similar pattern: 135 students (20.30%) scored high, 133 (20.00%) moderate, and 397 (59.70%) low. In Politics, 81 students (21.72%) achieved high marks, 73 (19.57%) moderate, and 219 (58.71%) low. History showed slightly better high-end performance with 93 students (25.00%) in the top group, followed by 68 (18.28%) moderate and 211 (56.72%) low. For Geography, 98 students (21.26%) were high performance, 91 (19.74%) moderate, and 272 (59.00%) low. The total score distribution revealed 229 students (20.32%) in the high range, 223 (19.79%) moderate, and 675 (59.89%) low.

Table 4.6: Examination Performance after Subject Options in 2024

Subjects	**The number of students (Percentage %)**		
	High	**Moderate**	**Low**
Chinese	256 (22.72)	211 (18.72)	660 (58.56)
Mathematics	231 (20.50)	220 (19.52)	676 (59.98)
English	225 (19.96)	217 (19.25)	685 (60.78)
Physics	152 (20.13)	149 (19.74)	454 (60.13)
Chemistry	152 (20.13)	153 (20.26)	450 (59.60)
Biology	135 (20.30)	133 (20.00)	397 (59.70)
Politics	81 (21.72)	73 (19.57)	219 (58.71)
History	93 (25.00)	68 (18.28)	211 (56.72)
Geography	98 (21.26)	91 (19.74)	272 (59.00)
Total scores	229 (20.32)	223 (19.79)	675 (59.89)

In the 2025 New College Entrance Examination, student performance showed a consistent trend across all measured subjects (see Table 4.7). Chinese results indicated that 222 students (19.70%) performed at a high level, 248 (22.01%) fell within the moderate range, and 657 (58.30%)

scored low. For Mathematics, the number of high scorers was also 222 (19.70%), with 234 students (20.76%) performing moderately and 671 (59.54%) scoring low. English followed a similar distribution: 222 students (19.70%) achieved high scores, 230 (20.41%) were moderate, and 675 (59.89%) scored low. Physics scores included 149 students (19.74%) in the high category, 156 (20.66%) moderate, and 450 (59.60%) low. In Chemistry, only 132 students (17.48%) were high achievers, 160 (21.19%) were moderate, and 463 (61.32%) fell into the low group. Biology outcomes comprised 132 students (19.85%) scoring high, 140 (21.05%) moderate, and 393 (59.10%) low. The number of high scorers in Politics was 73 (19.57%), with 68 moderate (18.23%) and 232 low (62.20%). History results showed 75 students (20.16%) at the high level, 70 (18.82%) moderate, and 227 (61.02%) low. In Geography, 90 students (19.52%) scored high, 91 (19.74%) moderate, and 280 (60.74%) low. The total score distribution recorded 227 high (20.14%), 225 moderate (19.96%), and 675 low (59.89%).

Table 4.7: NCEE Performance in 2025

Subjects	**The number of students (Percentage %)**		
	High	**Moderate**	**Low**
Chinese	222 (19.70)	248 (22.01)	657 (58.30)
Mathematics	222 (19.70)	234 (20.76)	671 (59.54)
English	222 (19.70)	230 (20.41)	675 (59.89)
Physics	149 (19.74)	156 (20.66)	450 (59.60)
Chemistry	132 (17.48)	160 (21.19)	463 (61.32)
Biology	132 (19.85)	140 (21.05)	393 (59.10)
Politics	73 (19.57)	68 (18.23)	232 (62.20)
History	75 (20.16)	70 (18.82)	227 (61.02)
Geography	90 (19.52)	91 (19.74)	280 (60.74)
Total scores	227 (20.14)	225 (19.96)	675 (59.89)

4.2.2 Family Dimensions

Table 4.8 shows the distribution across four family factors, which include father's education level, mother's education level, annual family income, and parental involvement in students' learning.

Table 4.8: Distribution of Levels across Four Family Factors

Items	The number of parents (Percentage %)		
	High	**Moderate**	**Low**
Father's education level	300 (26.62)	205 (18.19)	622 (55.19)
Mother's education level	310 (27.51)	227 (20.14)	590 (52.35)
Annual family income	291 (25.82)	219 (19.43)	617 (54.75)
Parental involvement in students' learning	265 (23.51)	222 (19.70)	640 (56.79)

The educational level of fathers had a measurable impact on students' academic performance. A total of 300 students (26.62%) were associated with fathers holding a high level of education. Meanwhile, 205 students (18.19%) belonged to the moderate category, and the majority 622 students (55.19%) had fathers with comparatively lower educational attainment.

Similarly, the mother's educational background exhibited a parallel trend. Among the sample, 310 students (27.51%) had mothers with high education levels. Moderate-level maternal education was recorded for 227 students (20.14%), and 590 students (52.35%) had mothers with lower educational attainment. The slightly higher percentage of mothers in the high-education group compared to fathers may reflect evolving trends in maternal influence on learning habits, especially in subjects requiring verbal reasoning and long-term planning.

Family income levels were also considered an influential variable. High-income families accounted for 291 students (25.82%), while 219 students (19.43%) came from moderate-income households. A significant portion 617 students (54.75%) belonged to families with lower annual income.

Lastly, parental involvement in students' learning emerged as a critical determinant. In this category, 265 students (23.51%) benefited from high levels of parental support, while 222 students (19.70%) experienced moderate involvement. Notably, 640 students (56.79%) received low levels of academic guidance from their parents.

4.2.3 School Dimensions

a) Teacher's education level, qualification, and self-efficacy

Table 4.9 shows the distribution across school factors, which include teacher's education level, teacher's qualification, and teacher's self-efficacy.

Table 4.9: Distribution of Teacher's Education Level, Qualification, and Self-efficacy

Items	**The number of teachers (Percentage %)**		
	High	**Moderate**	**Low**
Teacher's education level	30 (34.09)	58 (65.91)	0 (0)
Teacher's qualification	56 (63.64)	26 (29.55)	6 (6.82)
Teacher's self-efficacy	38 (43.18)	30 (34.09)	20 (22.73)

The education level of teachers was analysed. Out of the total respondents, 30 teachers (34.09%) possessed high-level education level, while 58 teachers (65.91%) were categorized under the moderate level. Notably, there were no teachers (0%) in the low education level category. This distribution suggests that most teachers held at least a moderate level of academic training, which may support the baseline quality of instruction delivered across schools.

With regard to teacher's qualification, 56 teachers (63.64%) were reported to hold high-level certifications. Another 26 teachers (29.55%) held moderate-level qualifications, while only 6 teachers (6.82%) were classified under the low qualification group. This data shows a strong trend toward professional accreditation among teaching staff, suggesting that a majority met or exceeded established qualification standards.

Self-efficacy among teachers was more evenly spread. A total of 38 teachers (43.18%) reported high self-efficacy in their teaching abilities. Meanwhile, 30 teachers (34.09%) fell into the moderate category, and 20 (22.73%) exhibited low levels of teaching confidence. These findings indicate that while a considerable number of educators feel competent in managing classroom demands and delivering effective instruction, a notable portion still struggles with confidence and instructional clarity. Given the influence of self-efficacy on instructional strategy and classroom management, these results suggest that improving teacher confidence could lead to measurable gains in student academic performance across multiple disciplines.

b) Teaching methods

Table 4.10 shows the numbers and proportions of the school factors affecting student performance related to teachers' teaching methods.

Table 4.10: Distribution of Teachers' Teaching Methods

Items	Attributes	Numbers	Proportions (%)
Teaching methods	Traditional teaching method	28	31.82
	Smart classroom teaching method	39	44.32
	Both	21	23.86

Table 4.10 shows that traditional teaching was used by 28 teachers, accounting for 31.82%

of the total. A larger group, comprising 39 teachers (44.32%), adopted smart classroom teaching methods, indicating a growing trend toward technology-enhanced instruction. In addition, 21 teachers (23.86%) reported using both traditional and smart classroom strategies in their daily practice. This mixed-method approach suggests an effort to integrate innovation with conventional techniques. The data imply that while modern tools are increasingly embraced, a significant portion of educators still relies on traditional models or seeks to balance both. These instructional patterns may influence classroom engagement, resource utilization, and ultimately, student learning outcomes.

4.2.4 Social Dimensions

a) Peer relationships, and socioeconomic status

Table 4.11 shows the distribution across two social factors, which are peer relationships and socioeconomic status.

Table 4.11: Distribution of Peer Relationships, and Socioeconomic Status

Items	The number of students (Percentage %)		
	High	Moderate	Low
Peer relationships	243 (21.56)	238 (21.12)	646 (57.32)
Socioeconomic status	289 (25.64)	431 (38.24)	407 (36.11)

Table 4.11 shows that a total of 243 students (21.56%) were categorized as having strong peer relationships, which were generally associated with improved classroom engagement, collaborative learning, and emotional support. Another 238 students (21.12%) maintained moderate peer interactions. The majority 646 students (57.32%) fell into the low category, indicating limited social connection or the presence of negative peer influences. The high proportion of students with weak peer relationships raises concerns, as such conditions can contribute to reduced academic motivation, increased absenteeism, and a decline in overall academic performance, especially in

group-based or discussion-heavy subjects.

Socioeconomic status (SES) presented a more even distribution. Among the respondents, 289 students (25.64%) belonged to families with high SES. A larger group, comprising 431 students (38.24%), fell under the moderate SES classification. Meanwhile, 407 students (36.11%) were identified as having low socioeconomic backgrounds. Students from high-SES families generally benefitted from better access to educational resources, enrichment programs, and private tutoring. In contrast, those from low-SES backgrounds often faced constraints such as limited study materials, financial stress, or less parental availability. The relatively balanced proportions across SES levels suggest a wide socioeconomic span within the student population, which may lead to performance inequality and differential educational trajectories over time.

b) Social support, and private tutoring

Table 4.12 shows the distribution of social factors related to social support, and private tutoring.

Table 4.12: Distribution of Social Support, and Private Tutoring

Items	Attributes	Numbers	Proportions (%)
Social support	Yes	102	9.05
	No	1025	90.95
Private tutoring	Yes	61	5.41
	No	1066	94.59

The results from Table 4.12 show that only 102 students (9.05%) reported having consistent access to social support, including emotional encouragement, academic assistance, or mentorship from family members, peers, or school staff. In contrast, a vast majority of 1,025 students (90.95%) indicated that they lacked sufficient social support in their learning environment. This discrepancy highlights a potential area of concern, as the absence of supportive relationships is often linked to increased stress levels, reduced academic resilience, and diminished engagement in school activities. The extremely low proportion of socially supported students suggests that the emotional and interpersonal dimensions of student development may be under-addressed in the current educational context.

Private tutoring also appeared to be limited in reach. Only 61 students (5.41%) participated in private tutoring sessions, while 1,066 students (94.59%) did not receive any form of academic instruction outside of regular school hours. The low rate of tutoring participation implies either restricted access due to economic barriers or a lack of tutoring culture within the studied population. Private tutoring is widely recognized for its role in addressing learning gaps, reinforcing classroom content, and enhancing examination readiness. Therefore, the overwhelming absence of such supplementary learning opportunities may contribute to persistent performance disparities, particularly among students requiring individualized support.

4.3 Data Preparation and Preprocessing

Data preparation and preprocessing play essential roles in ensuring the quality and reliability of datasets used for analysis. Preprocessing includes tasks such as cleaning, normalization, handling missing values, and transforming data to make it ready for statistical or machine learning applications. In this study, the dataset initially included 1,289 students and 103 teachers. However, due to missing data and the refusal of some participants, data from only 1,127 students and 88 teachers were analysed. Missing data can introduce bias and reduce the dataset's representativeness, necessitating careful handling using imputation or deletion techniques (Bala & Behal, 2024). By cleaning and transforming data, preprocessing enhances its consistency and relevance for further analysis.

The importance of preprocessing is further highlighted by the need to address challenges such as noise, outliers, and imbalances in the dataset. For instance, feature selection techniques can help identify and retain significant variables while reducing redundancy (Dol & Jawandhiya, 2024). In this context, attributes with limited variability or high correlation were excluded to simplify the dataset without compromising its integrity. These preprocessing steps are critical to minimize distortions and ensure the dataset's suitability for data transformation and subsequent analytical stages. By addressing the missing responses and inconsistencies, the refined dataset is prepared for further analysis, enhancing the validity of the study's conclusions.

4.4 Data Transformation

Data transformation is the process of changing raw data from nominal data to numerical data for analysis. This is important for educational data mining, especially for prediction and classification tasks. It helps make data usable for machine learning models. For example, student performance data from different sources, such as past performance, motivation level, and self-efficacy, can be turned into numbers. These numbers help algorithms find patterns and trends. One study showed how data transformation using ICT tools improved teaching and learning by analysing digital data from online classes (Ashish & Anitha, 2024). Another study used real-life projects to guide students in transforming educational data into numerical formats. This method supported skill-based learning and linked education to job market needs (Wang et al., 2024). Data transformation involves steps like normalizing values, handling missing data, and coding categorical data into numerical formats. For instance, text labels like "Low", "Moderate" and "High" can be replaced with numbers like 1, 2, and 3. Transforming data allows systems to predict student performance. These methods ensure consistency in data analysis and improve the accuracy of predictive models. By preparing data properly, researchers and educators can uncover valuable insights that enhance learning experiences and support decision-making. Table 4.13 shows that all current data has been transformed into numerical data. Nominal data is transformed into numerical data through encoding techniques such as label encoding or one-hot encoding, enabling its processing by computational algorithms.

Table 4.13: Nominal Data Transformed into Numerical Data

Items	Nominal data	Numerical data
Age	14 years old	1
	15 years old	2
	16 years old	3
	17 years old	4
	18 years old	5
	20 years old	6
Gender	Female	1
	Male	2
Motivation level	Low	1
	Moderate	2
	High	3
Self-efficacy	Low	1
	Moderate	2
	High	3
Test anxiety	Low	1
	Moderate	2
	High	3
Past performance	Low: Below 60% of aggregate score	1
	Moderate: More than 60% of aggregate score (including 60%) and less than 80%	2
	High: More than 80% of aggregate score (including 80%)	3
Father's education level	Low: Junior high school and below, and senior high school	1
	Moderate: Undergraduate degree	2
	High: Master's degree and doctoral degree	3
Mother's education level	Low: Junior high school and below, and senior high school	1

Table 4.13 continued

	Moderate: Undergraduate degree	2
	High: Master's degree and doctoral degree	3
Family income	Low	1
	Moderate	2
	High	3
Parental involvement in student's learning	Low	1
	Moderate	2
	High	3
Teacher's education level	Low: Undergraduate degree	1
	Moderate: Undergraduate degree	2
	High: Master's degree and doctoral degree	3
Teacher's qualification	Low: Secondary school teacher II	1
	Moderate: Secondary school teacher I	2
	High: Secondary school associate senior teacher, and secondary school full senior teacher	3
Teaching methods	Traditional teaching method	1
	Smart classroom teaching methods	2
	Both	3
Teacher's self-efficacy	Low	1
	Moderate	2
	High	3
Peer relationships	Low	1
	Moderate	2
	High	3
Social support	Yes	1
	No	2
Private tutoring	Yes	1

Table 4.13 continued

	No	2
Socioeconomic status	Low	1
	Moderate	2
	High	3

4.5 Identification the Most Important Factors to Predict Student NCEE Performance in Chinese, Mathematics and English

This section shows the results of the most important factors to predict student NCEE performance from Table 4.14. It looks at Chinese, Mathematics, and English subjects. Different algorithms are used to find key factors for each subject. The algorithms are Naïve Bayes (NB), Decision Tree (DT), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs). For each algorithm, the section shows the process and the important factors. It also gives the weights of each factor. Key factors include teaching method, teacher's self-efficacy, teacher's qualification, socioeconomic status, gender, 4 English, parental involvement in student's learning, mother's education level, 4 Chinese, and peer relationship. Each algorithm gives different weights to these factors. This helps to find the most important factors for each subject. Educators and policymakers can use this information to improve teaching and support for students.

4.5.1 The Most Important Factors to Predict Student NCEE Performance in Chinese Subject

RQ1a: What are the most important factors to predict student NCEE performance in Chinese subject based on NB, DT, ANNs, SVMs algorithms?

a) NB

Figure 4.1 shows the process of identifying the most important factors to predict student

NCEE performance in the Chinese subject using NB.

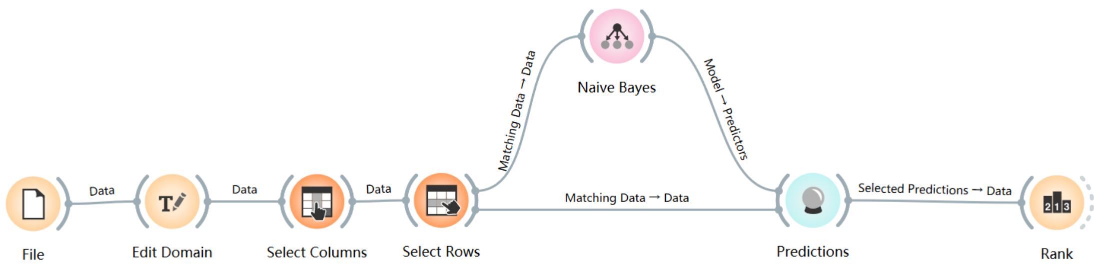

Figure 4.1: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the Chinese Subject Using NB

Table 4.14 lists the weight of attributes for predicting student NCEE performance in Chinese subject using NB. Each "Weight" represents the significance of each attribute in the prediction.

Table 4.14: Weight of the Attributes for Predicting Student NCEE Performance in Chinese Subject Using NB

Attributes for Chinese	Weight
Motivation level	0.845
Teaching method	0.110
Gender	0.094
4 Chinese	0.087
Teacher's self-efficacy	0.076
4 English	0.068
3 Chinese	0.065
2 Politics	0.056
2 Biology	0.056
Teacher's qualification	0.052
Parental involvement in student's learning	0.051

Table 4.14 continued

1 Physics	0.050
3 Geography	0.048
1 Chinese	0.042
Father's education level	0.038
3 Politics	0.037
2 Physics	0.036
2 Geography	0.035
3 Physics	0.033
Socioeconomic status	0.031
2 Mathematics	0.029
Peer relationship	0.027
4 Total scores	0.027
2 English	0.027
2 Chinese	0.025
Test anxiety	0.025
3 Total scores	0.023
3 Chemistry	0.023
Student's self-efficacy	0.021
2 Total scores	0.021
1 History	0.020
Teacher's education level	0.019
4 Mathematics	0.015
1 Mathematics	0.014
Social support	0.012
Annual family income	0.010
1 English	0.009
2 History	0.007
3 English	0.007

Table 4.14 continued

3 Biology	0.005

The initial prediction accuracy rate of Chinese performance using NB was 68.1%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.15, continuing until the highest prediction accuracy is achieved.

Table 4.15: Factors Removed from Predicting Student NCEE Performance in Chinese Subject Using NB

No.	Factors Removed	Prediction accuracy (%)
1	3 Biology	68.1
2	3 Biology, 3 English	68.1
3	3 Biology, 3 English, 2 History	68.1
4	3 Biology, 3 English, 2 History, 1 English	70.6
5	3 Biology, 3 English, 2 History, 1 English, Annual family income	71.2
6	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support	71.2
7	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics	72.1
8	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics	72.1
9	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level	75.8
10	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History	75.8
11	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores	75.8

Table 4.15 continued

12	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy	75.8
13	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry	75.9
14	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores	77.8
15	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety	77.8
16	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese	80.0
17	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English	80.0
18	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores	80.0
19	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship	81.1
20	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics	81.1
21	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status	85.3
22	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status, 3 Physics	87.1

Table 4.15 continued

23	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status, 3 Physics, 2 Geography	87.1
24	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status, 3 Physics, 2 Geography, 2 Physics	87.7
25	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status, 3 Physics, 2 Geography, 2 Physics, 3 Politics	91.5
26	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status, 3 Physics, 2 Geography, 2 Physics, 3 Politics, Father's education level	91.5
27	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status, 3 Physics, 2 Geography, 2 Physics, 3 Politics, Father's education level, 1 Chinese	91.5
28	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status, 3 Physics, 2 Geography, 2 Physics, 3 Politics, Father's education level, 1 Chinese, 3 Geography	95.1
29	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status, 3 Physics, 2 Geography, 2 Physics, 3 Politics, Father's education level, 1 Chinese, 3 Geography, 1 Physics	95.1

Table 4.15 continued

30	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status, 3 Physics, 2 Geography, 2 Physics, 3 Politics, Father's education level, 1 Chinese, 3 Geography, 1 Physics, Parental involvement in student's learning	95.1
31	3 Biology, 3 English, 2 History, 1 English, Annual family income, Social support, 1 Mathematics, 4 Mathematics, Teacher's education level, 1 History, 2 Total scores, Student's self-efficacy, 3 Chemistry, 3 Total scores, Test anxiety, 2 Chinese, 2 English, 4 Total scores, Peer relationship, 2 Mathematics, Socioeconomic status, 3 Physics, 2 Geography, 2 Physics, 3 Politics, Father's education level, 1 Chinese, 3 Geography, 1 Physics, Parental involvement in student's learning, Teacher's qualification	86.2

According to Table 4.15, when factors, such as 3 Biology, 3 English, 2 History, 1 English, annual family income, social support, 1 Mathematics, 4 Mathematics, teacher's education level, 1 History, 2 total scores, student's self-efficacy, 3 Chemistry, 3 total scores, test anxiety, 2 Chinese, 2 English, 4 total scores, peer relationship, 2 Mathematics, socioeconomic status, 3 Physics, 2 Geography, 2 Physics, 3 Politics, father's education level, 1 Chinese, 3 Geography, 1 Physics, and parental involvement in student's learning are removed, the prediction accuracy reaches 95.1%, while when factor teacher's qualification is removed, the prediction accuracy drops to 86.2%. From this, it can be known that factors such as motivation level, teaching method, gender, 4 Chinese, teacher's self-efficacy, 4 English, 3 Chinese, 2 Politics, 2 Biology, and teacher's qualification are suitable for predicting student NCEE performance in Chinese subject using NB.

b) DT

Figure 4.2 shows the process of identifying the most important factors to predict student NCEE performance in Chinese subject using DT.

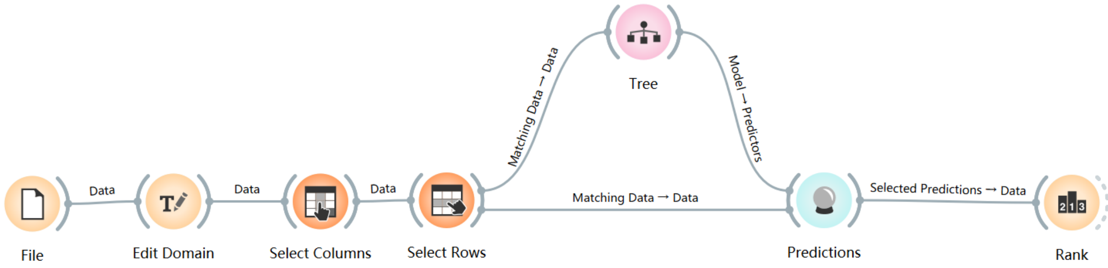

Figure 4.2: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the Chinese Subject Using DT

Table 4.16 lists the weight of attributes for predicting student NCEE performance in Chinese subject using DT.

Table 4.16: Weight of the Attributes for Predicting Student NCEE Performance in Chinese Subject Using DT

Attributes for Chinese	Weight
Motivation level	0.870
Teaching method	0.103
Teacher's qualification	0.095
4 English	0.090
Annual family income	0.088
Parental involvement in student's learning	0.084
Gender	0.075
Teacher's education level	0.073
3 Physics	0.066
4 Chinese	0.064
Father's education level	0.058
3 Physics	0.047
4 Chinese	0.042

Table 4.16 continued

Father's education level	0.041
Mother's education level	0.039
3 Chemistry	0.037
3 Total scores	0.036
3 Politics	0.036
1 History	0.036
1 Politics	0.034
4 Mathematics	0.031
2 English	0.030
2 Mathematics	0.026
2 Chinese	0.026
1 Chinese	0.022
1 English	0.019
4 Total scores	0.017
1 Chemistry	0.015
1 Physics	0.015
2 Geography	0.013
3 Biology	0.011
3 History	0.011
3 Geography	0.010
Private tutoring	0.009
Test anxiety	0.009
3 English	0.008
2 History	0.008
Peer relationship	0.008
Student's self-efficacy	0.007
2 Physics	0.007
Social support	0.006

Table 4.16 continued

Age	0.005
1 Mathematics	0.002
2 Chemistry	0.002

The initial prediction accuracy rate of Chinese performance using DT was 72.5%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.17, continuing until the highest prediction accuracy is achieved.

Table 4.17: Factors Removed from Predicting Student NCEE Performance in Chinese Subject Using DT

No.	Factors Removed	Prediction accuracy (%)
1	2 Chemistry	72.5
2	2 Chemistry, 1 Mathematics	74.7
3	2 Chemistry, 1 Mathematics, Age	74.7
4	2 Chemistry, 1 Mathematics, Age, Social support	74.7
5	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics	76.0
6	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy	76.0
7	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship	76.4
8	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History	76.4
9	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English	76.4
10	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety	77.0
11	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring	80.9

Table 4.17 continued

12	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography	80.9
13	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History	83.7
14	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology	83.7
15	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography	83.7
16	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics	84.9
17	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry	84.9
18	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores	86.3
19	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English	87.2
20	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese	87.2
21	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese	88.1
22	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics	88.1

Table 4.17 continued

23	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English	88.1
24	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English, 4 Mathematics	88.1
25	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English, 4 Mathematics, 1 Politics	90.8
26	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English, 4 Mathematics, 1 Politics, 1 History	90.8
27	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English, 4 Mathematics, 1 Politics, 1 History, 3 Politics	90.8
28	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English, 4 Mathematics, 1 Politics, 1 History, 3 Politics, 3 Total scores	91.6
29	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English, 4 Mathematics, 1 Politics, 1 History, 3 Politics, 3 Total scores, 3 Chemistry	92.9
30	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English, 4 Mathematics, 1 Politics, 1 History, 3 Politics, 3 Total scores, 3 Chemistry, Mother's education level	92.9

Table 4.17 continued

31	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English, 4 Mathematics, 1 Politics, 1 History, 3 Politics, 3 Total scores, 3 Chemistry, Mother's education level, Father's education level	92.9
32	2 Chemistry, 1 Mathematics, Age, Social support, 2 Physics, Student's self-efficacy, Peer relationship, 2 History, 3 English, Test anxiety, Private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 Total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English, 4 Mathematics, 1 Politics, 1 History, 3 Politics, 3 Total scores, 3 Chemistry, Mother's education level, Father's education level, 4 Chinese	88.7

According to Table 4.17, when factors, such as 2 Chemistry, 1 Mathematics, age, social support, 2 Physics, student's self-efficacy, peer relationship, 2 History, 3 English, test anxiety, private tutoring, 3 Geography, 3 History, 3 Biology, 2 Geography, 1 Physics, 1 Chemistry, 4 total scores, 1 English, 1 Chinese, 2 Chinese, 2 Mathematics, 2 English, 4 Mathematics, 1 Politics, 1 History, 3 Politics, 3 total scores, 3 Chemistry, mother's education level, and father's education level are removed, the prediction accuracy reaches 92.9%, while when 4 Chinese is removed, the prediction accuracy drops to 88.7%. From this, it can be known that factors such as motivation level, teaching method, teacher's qualification, 4 English, annual family income, parental involvement in student's learning, gender, teacher's self-efficacy, 3 Chinese, socioeconomic status, teacher's education level, 3 Physics, and 4 Chinese are suitable for predicting student NCEE performance in Chinese subject using DT.

c) ANNs

Figure 4.3 shows the process of identifying the most important factors to predict student NCEE performance in Chinese subject using ANNs.

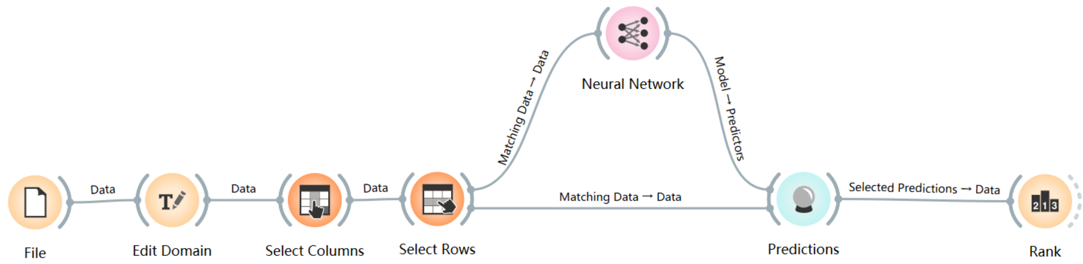

Figure 4.3: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the Chinese Subject Using ANNs

Table 4.18 lists the weight of attributes for predicting student NCEE performance in Chinese subject using ANNs.

Table 4.18: Weight of the Attributes for Predicting Student NCEE Performance in Chinese Subject Using ANNs

Attributes for Chinese	**Weight**
Motivation level	0.820
3 Chinese	0.123
Teacher's qualification	0.089
Teacher's self-efficacy	0.071
Parental involvement in student's learning	0.062
2 Chinese	0.060
4 Chinese	0.060
Teacher's education level	0.058
Teaching method	0.055
Gender	0.051
Mother's education level	0.050
1 Mathematics	0.048
Peer relationship	0.046
3 Geography	0.043

Table 4.18 continued

3 Biology	0.035
Annual family income	0.024
2 History	0.024
2 Total scores	0.021
4 English	0.019
4 Mathematics	0.018
1 Physics	0.018
Age	0.015
Socioeconomic status	0.015
Private tutoring	0.012
1 Chinese	0.009
Test anxiety	0.008
2 Physics	0.006
3 Chemistry	0.005
2 Politics	0.005
2 Biology	0.005
2 Geography	0.004
Student's self-efficacy	0.003

The initial prediction accuracy rate of Chinese performance using ANNs was 78.4%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.19, continuing until the highest prediction accuracy is achieved.

Table 4.19: Factors Removed from Predicting Student NCEE Performance in Chinese Subject Using ANNs

No.	Factors Removed	Prediction accuracy (%)
1	Student's self-efficacy	78.4
2	Student's self-efficacy, 2 Geography	78.4
3	Student's self-efficacy, 2 Geography, 2 Biology	78.4
4	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics	81.9
5	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry	81.9
6	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics	86.8
7	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety	86.8
8	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese	86.8
9	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring	90.9
10	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status	90.9
11	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status, Age	91.7
12	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status, Age, 1 Physics	91.7
13	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status, Age, 1 Physics, 4 Mathematics	92.0
14	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status, Age, 1 Physics, 4 Mathematics, 4 English	92.0
15	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status, Age, 1 Physics, 4 Mathematics, 4 English, 2 Total scores	92.0
16	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status, Age, 1 Physics, 4 Mathematics, 4 English, 2 Total scores, 2 History	92.0

Table 4.19 continued

17	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status, Age, 1 Physics, 4 Mathematics, 4 English, 2 Total scores, 2 History, Annual family income	93.8
18	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status, Age, 1 Physics, 4 Mathematics, 4 English, 2 Total scores, 2 History, Annual family income, 3 Biology	93.8
19	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status, Age, 1 Physics, 4 Mathematics, 4 English, 2 Total scores, 2 History, Annual family income, 3 Biology, 3 Geography	93.8
20	Student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, Test anxiety, 1 Chinese, Private tutoring, Socioeconomic status, Age, 1 Physics, 4 Mathematics, 4 English, 2 Total scores, 2 History, Annual family income, 3 Biology, 3 Geography, Peer relationship	85.8

According to Table 4.19, when factors, such as student's self-efficacy, 2 Geography, 2 Biology, 2 Politics, 3 Chemistry, 2 Physics, test anxiety, 1 Chinese, private tutoring, socioeconomic status, age, 1 Physics, 4 Mathematics, 4 English, 2 total scores, 2 History, annual family income, 3 Biology, and 3 Geography are removed, the prediction accuracy reaches 93.8%, while when peer relationship is removed, the prediction accuracy drops to 85.8%. From this, it can be known that factors such as motivation level, 3 Chinese, teacher's qualification, teacher's self-efficacy, parental involvement in student's learning, 2 Chinese, 4 Chinese, teacher's education level, teaching method, gender, mother's education level, 1 Mathematics and peer relationship are suitable for predicting student NCEE performance in Chinese subject using ANNs.

d) SVMs

Figure 4.4 shows the process of identifying the most important factors to predict student NCEE performance in Chinese subject using SVMs.

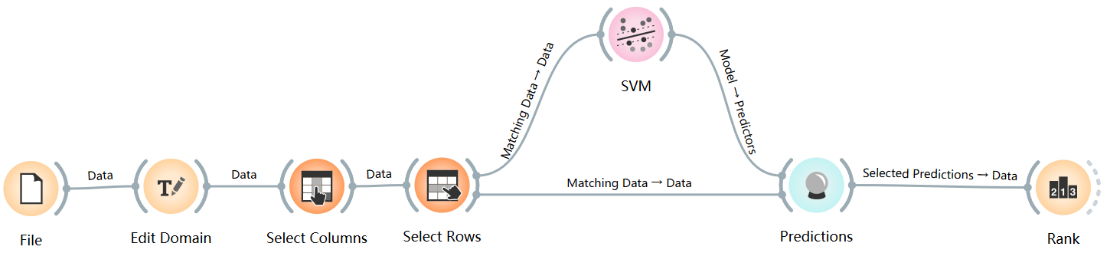

Figure 4.4: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the Chinese Subject Using SVMs

Table 4.20 lists the weight of attributes for predicting student NCEE performance in Chinese subject using SVMs.

Table 4.20: Weight of the Attributes for Predicting Student NCEE Performance in Chinese Subject Using SVMs

Attributes for Chinese	Weight
Motivation level	0.806
Teaching method	0.135
Teacher's self-efficacy	0.130
Teacher's qualification	0.091
Age	0.082
4 English	0.080
2 Chinese	0.066
3 Chinese	0.065
Parental involvement in student's learning	0.065
1 Total scores	0.057
Teacher's education level	0.048
1 English	0.047
Annual family income	0.046
3 History	0.039

Table 4.20 continued

1 Politics	0.039
3 Biology	0.035
3 Geography	0.032
1 Chinese	0.030
1 Physics	0.030
Peer relationship	0.029
Father's education level	0.027
2 History	0.027
Gender	0.027
Socioeconomic status	0.026
Student's self-efficacy	0.025
2 English	0.021
4 Chinese	0.019
3 Mathematics	0.018
Mother's education level	0.018
Test anxiety	0.016
Private tutoring	0.016
1 Mathematics	0.016
3 Chemistry	0.015
3 English	0.013
3 Politics	0.009
2 Physics	0.009
4 Mathematics	0.009
2 Politics	0.005
2 Biology	0.005
3 Physics	0.004
1 Chemistry	0.003
1 History	0.002

The initial prediction accuracy rate of Chinese performance using SVMs was 69.8%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.21, continuing until the highest prediction accuracy is achieved.

Table 4.21: Factors Removed from Predicting Student NCEE Performance in Chinese Subject Using SVMs

No.	Factors Removed	Prediction accuracy (%)
1	1 History	69.8
2	1 History, 1 Chemistry	69.8
3	1 History, 1 Chemistry, 3 Physics	72.1
4	1 History, 1 Chemistry, 3 Physics, 2 Biology	72.1
5	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics	72.1
6	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics	74.4
7	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics	74.4
8	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics	75.1
9	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English	75.1
10	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry	77.7
11	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics	77.7
12	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring	77.7
13	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety	79.0

Table 4.21 continued

14	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level	79.0
15	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics	80.6
16	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese	80.6
17	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English	80.6
18	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy	80.6
19	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status	84.4
20	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender	84.4
21	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History	86.6
22	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History, Father's education level	86.8
23	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History, Father's education level, Peer relationship	86.8

Table 4.21 continued

24	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History, Father's education level, Peer relationship, 1 Physics	86.8
25	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History, Father's education level, Peer relationship, 1 Physics,1 Chinese	90.1
26	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History, Father's education level, Peer relationship, 1 Physics,1 Chinese, 3 Geography	90.1
27	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History, Father's education level, Peer relationship, 1 Physics,1 Chinese, 3 Geography, 3 Biology	92.9
28	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History, Father's education level, Peer relationship, 1 Physics,1 Chinese, 3 Geography, 3 Biology, 1 Politics	92.9
29	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History, Father's education level, Peer relationship, 1 Physics,1 Chinese, 3 Geography, 3 Biology, 1 Politics, 3 History	92.9
30	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History, Father's education level, Peer relationship, 1 Physics,1 Chinese, 3 Geography, 3 Biology, 1 Politics, 3 History, Annual family income	92.9

Table 4.21 continued

31	1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, Private tutoring, Test anxiety, Mother's education level, 3 Mathematics, 4 Chinese, 2 English, Student's self-efficacy, Socioeconomic status, Gender, 2 History, Father's education level, Peer relationship, 1 Physics,1 Chinese, 3 Geography, 3 Biology, 1 Politics, 3 History, Annual family income, 1 English	84.4

According to Table 4.21, when factors, such as 1 History, 1 Chemistry, 3 Physics, 2 Biology, 2 Politics, 4 Mathematics, 2 Physics, 3 Politics, 3 English, 3 Chemistry, 1 Mathematics, private tutoring, test anxiety, mother's education level, 3 Mathematics, 4 Chinese, 2 English, student's self-efficacy, socioeconomic status, gender, 2 History, father's education level, peer relationship, 1 Physics,1 Chinese, 3 Geography, 3 Biology, 1 Politics, 3 History, and annual family income are removed, the prediction accuracy reaches 92.9%, while when 1 English is removed, the prediction accuracy drops to 84.4%. From this, it can be known that factors such as motivation level, teaching method, teacher's self-efficacy, teacher's qualification, age, 4 English, 2 Chinese, 3 Chinese, parental involvement in student's learning, 1 total scores, teacher's education level, and 1 English are suitable for predicting student NCEE performance in Chinese subject using SVMs.

In order to obtain the potentially best predictor factors of NCEE performance in Chinese, researcher temporarily take the top five affecting factors of each algorithm and then use four algorithms to make a re-prediction and check the prediction accuracy. By aggregating the top five predictive factors of each of the four algorithms and then eliminating the duplicate parts, four new sets of predictive factors and weight can be obtained (see Table 4.22), and also the prediction accuracy (see Table 4.23).

Table 4.22: Four New Sets of Predictive Factors and Weight to Repredict Student NCEE Performance in Chinese Subject using Four Algorithms

Factors using NB	Weight	Factors using DT	Weight	Factors using ANNs	Weight	Factors using SVMs	Weight

Table 4.22 continued

Motivation level	0.834	Motivation level	0.802	Motivation level	0.846	Motivation level	0.850
3 Chinese	0.127	Age	0.064	4 Chinese	0.110	Gender	0.072
Teacher's qualification	0.067	Gender	0.044	3 Chinese	0.107	Age	0.054
Parental involvement in student's learning	0.055	Annual family income	0.041	Gender	0.078	Teacher's self-efficacy	0.048
4 Chinese	0.052	4 Chinese	0.040	Teacher's qualification	0.067	Teaching method	0.041
Age	0.051	Parental involvement in student's learning	0.033	Annual family income	0.062	Teacher's qualification	0.040
Gender	0.042	Teacher's qualification	0.033	Teacher's self-efficacy	0.048	4 Chinese	0.028
Teacher's self-efficacy	0.042	3 Chinese	0.032	4 English	0.032	Parental involvement in student's learning	0.025
Teaching method	0.035	Teaching method	0.028	Parental involvement in student's learning	0.029	3 Chinese	0.024
Annual family income	0.007	Teacher's self-efficacy	0.010	Teaching method	0.029	Annual family income	0.021
4 English	0.002	4 English	0.006	Age	0.022	4 English	0.008

Table 4.23: The Prediction Accuracy of Repredicting Student NCEE Performance in Chinese Subject using Four Algorithms

Algorithm	Prediction Accuracy (%)	Algorithm	Prediction Accuracy (%)	Algorithm	Prediction Accuracy (%)	Algorithm	Prediction Accuracy (%)
NB	92.1	DT	90.9	ANNs	91.7	SVMs	89.7

As fine-tuning did not improve the prediction accuracy, the initial prediction using NB, which achieved the highest accuracy (95.1%), was used to determine the predictor factors. Referring to Table 4.14, the predictors of NCEE performance in Chinese subject are motivation level (weight 0.845) is the most critical factor, followed by teaching method (weight 0.110), gender (weight 0.094), 4 Chinese (weight 0.087), teacher's self-efficacy (weight 0.076), 4 English (weight 0.068), 3 Chinese (weight 0.065), 2 Politics (weight 0.056), 2 Biology (weight 0.056), and teacher's qualification (weight 0.052). Some of these predictors are also predictors of the other three algorithms (see Table 4.24).

Table 4.24: Same Predictors using NB in Other Three Algorithms in Chinese Subject

Factors using NB	Whether in DT	Whether in ANNs	Whether in SVMs
Motivation level	Yes	Yes	Yes
Teaching method	Yes	Yes	Yes
Gender	Yes	Yes	No
4 Chinese	Yes	Yes	No
Teacher's self-efficacy	Yes	Yes	Yes
4 English	Yes	No	Yes
3 Chinese	Yes	Yes	Yes
2 Politics	No	No	No
2 Biology	No	No	No
Teacher's qualification	Yes	Yes	Yes

4.5.2 The Most Important Factors to Predict Student NCEE Performance in Mathematics Subject

RQ1b: What are the most important factors to predict student NCEE performance in Mathematics subject based on NB, DT, ANNs, SVMs algorithms?

a) NB

Figure 4.5 shows the process of identifying the most important factors to predict student NCEE performance in the Mathematics subject using NB.

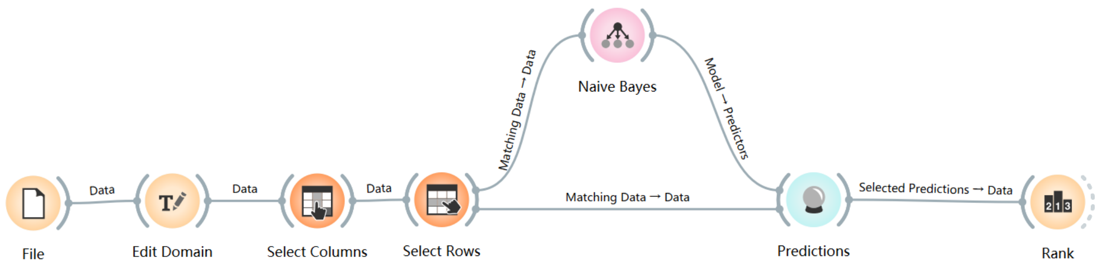

Figure 4.5: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the Mathematics Subject Using NB

Table 4.25 lists the weight of attributes for predicting student NCEE performance in the Mathematics subject using NB.

Table 4.25: Weight of the Attributes for Predicting Student NCEE Performance in Mathematics Subject Using NB

Attributes for Mathematics	Weight
Test anxiety	0.827
Father's education level	0.579
Mother's education level	0.543
Socioeconomic status	0.419
Peer relationship	0.391
Student's self-efficacy	0.344
2 Mathematics	0.192
Teaching method	0.134
4 Mathematics	0.132

Table 4.25 continued

3 Physics	0.123
Teacher' self-efficacy	0.118
3 Chemistry	0.112
4 Total scores	0.100
3 Mathematics	0.087
1 Mathematics	0.081
Teacher's qualification	0.078
4 English	0.075
1 English	0.071
2 Geography	0.069
Annual family income	0.064
Private tutoring	0.063
2 Physics	0.063
1 Physics	0.062
3 Biology	0.059
1 History	0.054
3 Geography	0.050
1 Politics	0.050
Teacher's education level	0.049
3 Chinese	0.047
Gender	0.046
4 Chinese	0.045
3 English	0.039
1 Chinese	0.036
2 Chemistry	0.035
2 English	0.033
2 Politics	0.028
2 Biology	0.028

Table 4.25 continued

3 Politics	0.028
2 History	0.024
Parental involvement in student's learning	0.024
Age	0.024
1 Total scores	0.020
3 History	0.020
2 Chinese	0.015
2 Total scores	0.010
Motivation level	0.007
Social support	0.005
1 Chemistry	0.004

The initial prediction accuracy rate of Mathematics performance using NB was 70.2%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.26, continuing until the highest prediction accuracy is achieved.

Table 4.26: Factors Removed from Predicting Student NCEE Performance in Mathematics Subject Using NB

No.	Factors Removed	Prediction accuracy (%)
1	1 Chemistry	70.2
2	1 Chemistry, Social support	70.2
3	1 Chemistry, Social support, Motivation level	70.2
4	1 Chemistry, Social support, Motivation level, 2 Total scores	71.0
5	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese	71.3

Table 4.26 continued

6	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History	71.3
7	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores	73.8
8	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age	73.8
9	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning	73.8
10	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History	75.0
11	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics	75.0
12	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology	75.0
13	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics	76.6
14	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English	76.6
15	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry	76.6
16	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese	76.6
17	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English	76.6
18	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese	81.9

Table 4.26 continued

19	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender	81.9
20	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese	81.9
21	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level	84.0
22	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics	85.1
23	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography	85.1
24	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History	85.1
25	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology	86.6
26	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics	86.6
27	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics	86.6

Table 4.26 continued

28	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics, Private tutoring	88.7
29	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics, Private tutoring, Annual family income	88.7
30	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics, Private tutoring, Annual family income, 2 Geography	92.3
31	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics, Private tutoring, Annual family income, 2 Geography, 1 English	92.3
32	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics, Private tutoring, Annual family income, 2 Geography, 1 English, 4 English	92.3
33	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics, Private tutoring, Annual family income, 2 Geography, 1 English, 4 English, Teacher's qualification	92.3
34	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics, Private tutoring, Annual family income, 2 Geography, 1 English, 4 English, Teacher's qualification, 1 Mathematics	94.6

Table 4.26 continued

35	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics, Private tutoring, Annual family income, 2 Geography, 1 English, 4 English, Teacher's qualification, 1 Mathematics, 3 Mathematics	96.4
36	1 Chemistry, Social support, Motivation level, 2 Total scores, 2 Chinese, 3 History, 1 Total scores, Age, Parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, Teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics, Private tutoring, Annual family income, 2 Geography, 1 English, 4 English, Teacher's qualification, 1 Mathematics, 3 Mathematics, 4 Total scores	87.7

According to Table 4.26, when factors, such as 1 Chemistry, social support, motivation level, 2 total scores, 2 Chinese, 3 History, 1 total scores, age, parental involvement in student's learning, 2 History, 3 Politics, 2 Biology, 2 Politics, 2 English, 2 Chemistry, 1 Chinese, 3 English, 4 Chinese, Gender, 3 Chinese, teacher's education level, 1 Politics, 3 Geography, 1 History, 3 Biology, 1 Physics, 2 Physics, private tutoring, annual family income, 2 Geography, 1 English, 4 English, teacher's qualification, 1 Mathematics, and 3 Mathematics are removed, the prediction accuracy reaches 96.4%, while when 4 total scores is removed, the prediction accuracy drops to 87.7%. From this, it can be known that factors such as test anxiety, father's education level, mother's education level, socioeconomic status, peer relationship, student's self-efficacy, 2 Mathematics, teaching method, 4 Mathematics, 3 Physics, teacher's self-efficacy, 3 Chemistry, and 4 total scores are suitable for predicting student NCEE performance in Mathematics subject using NB.

b) DT

Figure 4.6 shows the process of identifying the most important factors to predict student NCEE performance in the Mathematics subject using DT.

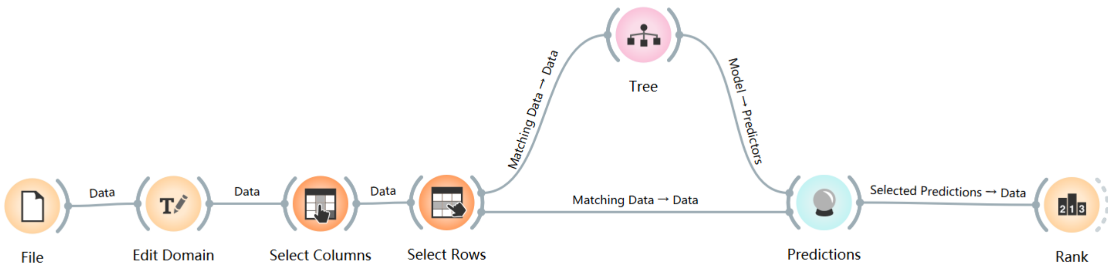

Figure 4.6: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the Mathematics Subject Using DT

Table 4.27 lists the weight of attributes for predicting student NCEE performance in the Mathematics subject using DT.

Table 4.27: Weight of the Attributes for Predicting Student NCEE Performance in Mathematics Subject Using DT

Attributes for Mathematics	Weight
Test anxiety	0.778
Father's education level	0.585
Mother' education level	0.504
Socioeconomic status	0.425
Student's self-efficacy	0.266
Peer relationship	0.250
4 Mathematics	0.196
2 Mathematics	0.181
Teacher's self-efficacy	0.159
Teaching method	0.150
3 Mathematics	0.144
3 Physics	0.141
1 Mathematics	0.123
Gender	0.123

Table 4.27 continued

Teacher's qualification	0.121
Teacher's education level	0.108
3 Chemistry	0.106
1 History	0.089
2 English	0.088
4 Total scores	0.088
Annual family income	0.076
2 Chemistry	0.075
2 Physics	0.072
3 Chinese	0.072
Parental involvement in student's learning	0.062
1 Politics	0.059
2 Chinese	0.056
Age	0.056
1 English	0.055
1 Physics	0.054
3 English	0.048
3 Geography	0.043
1 Chinese	0.040
Private tutoring	0.038
3 Biology	0.037
4 Chinese	0.029
2 Geography	0.028
2 Politics	0.028
2 Biology	0.028
3 Total scores	0.024
Motivation level	0.024

Table 4.27 continued

3 Politics	0.018
2 History	0.017
4 English	0.015
1 Chemistry	0.014
3 History	0.008

The initial prediction accuracy rate of Mathematics performance using DT was 71.0%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.28, continuing until the highest prediction accuracy is achieved.

Table 4.28: Factors Removed from Predicting Student NCEE Performance in Mathematics Subject Using DT

No.	Factors Removed	Prediction accuracy (%)
1	3 History	71.0
2	3 History, 1 Chemistry	72.9
3	3 History, 1 Chemistry, 4 English	72.9
4	3 History, 1 Chemistry, 4 English, 2 History	72.9
5	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics	74.9
6	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level	74.9
7	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores	74.9
8	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology	75.1
9	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics	75.1
10	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography	78.9

Table 4.28 continued

11	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese	78.9
12	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology	78.9
13	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring	80.7
14	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese	80.7
15	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography	83.9
16	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English	83.9
17	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics	83.9
18	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English	83.9
19	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English, Age	84.2
20	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English, Age, 2 Chinese	85.9
21	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English, Age, 2 Chinese, 1 Politics, Parental involvement in student’s learning, 3 Chinese, 2 Physics	85.9
22	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English, Age, 2 Chinese, 1 Politics, Parental involvement in student’s learning, 3 Chinese, 2 Physics, 2 Chemistry	89.7

Table 4.28 continued

23	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English, Age, 2 Chinese, 1 Politics, Parental involvement in student's learning, 3 Chinese, 2 Physics, 2 Chemistry, Annual family income	89.7
24	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English, Age, 2 Chinese, 1 Politics, Parental involvement in student's learning, 3 Chinese, 2 Physics, 2 Chemistry, Annual family income, 4 Total scores	89.7
25	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English, Age, 2 Chinese, 1 Politics, Parental involvement in student's learning, 3 Chinese, 2 Physics, 2 Chemistry, Annual family income, 4 Total scores, 2 English	92.9
26	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English, Age, 2 Chinese, 1 Politics, Parental involvement in student's learning, 3 Chinese, 2 Physics, 2 Chemistry, Annual family income, 4 Total scores, 2 English, 1 History	92.9
27	3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, Motivation level, 3 Total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, Private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English, Age, 2 Chinese, 1 Politics, Parental involvement in student's learning, 3 Chinese, 2 Physics, 2 Chemistry, Annual family income, 4 Total scores, 2 English, 1 History, 3 Chemistry	84.4

According to Table 4.28, when factors, such as 3 History, 1 Chemistry, 4 English, 2 History, 3 Politics, motivation level, 3 total scores, 2 Biology, 2 Politics, 2 Geography, 4 Chinese, 3 Biology, private tutoring, 1 Chinese, 3 Geography, 3 English, 1 Physics, 1 English, Age, 2 Chinese, 1 Politics, parental involvement in student's learning, 3 Chinese, 2 Physics, 2 Chemistry, annual family income, 4 total scores, 2 English, 1 History are removed, the prediction accuracy reaches 92.9%, while when 3 Chemistry is removed, the prediction accuracy drops to 84.4%. From this, it can be known that factors such as test anxiety, father's education level, mother's education level, socioeconomic status, student's self-efficacy, peer relationship, 4 Mathematics, 2 Mathematics, teacher's self-efficacy, teaching method, 3 Mathematics, 3 Physics, 1 Mathematics, gender, teacher's qualification, teacher's education level, and 3 Chemistry are suitable for predicting student NCEE performance in

Mathematics subject using DT.

c) ANNs

Figure 4.7 shows the process of identifying the most important factors to predict student NCEE performance in the Mathematics subject using ANNs.

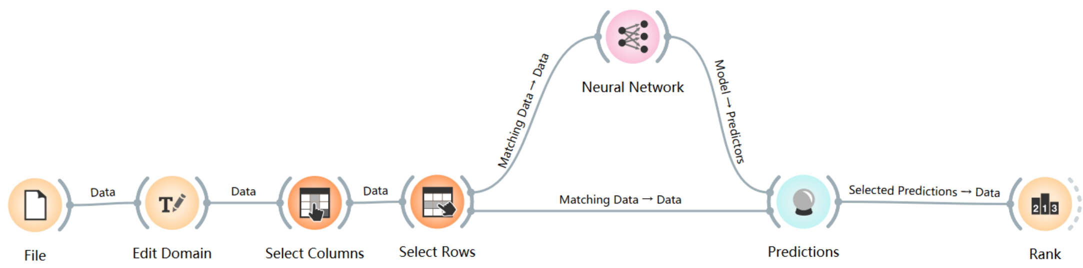

Figure 4.7: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the Mathematics Subject Using ANNs

Table 4.29 lists the weight of attributes for predicting student NCEE performance in the Mathematics subject using ANNs.

Table 4.29: Weight of the Attributes for Predicting Student NCEE Performance in Mathematics Subject Using ANNs

Attributes for Mathematics	Weight
Test anxiety	0.886
Mother’s education level	0.596
Father’s education level	0.488
Socioeconomic status	0.380
Peer relationship	0.301
Student’s self-efficacy	0.269

Table 4.29 continued

Teacher’s self-efficacy	0.189
Teaching method	0.184
4 Mathematics	0.144
3 Physics	0.122
2 Mathematics	0.117
3 Mathematics	0.108
Gender	0.101
Teacher’s qualification	0.098
Annual family income	0.074
3 Politics	0.068
3 Chinese	0.065
1 Chemistry	0.063
3 History	0.062
Teacher’s education level	0.060
3 Total scores	0.060
2 Chemistry	0.058
Age	0.058
2 Physics	0.057
1 History	0.057
1 Physics	0.052
1 Chinese	0.050
1 Total scores	0.049
3 Biology	0.048
4 Total scores	0.048
3 Geography	0.047
1 Mathematics	0.046
3 Chemistry	0.040
2 English	0.040

Table 4.29 continued

Parental involvement in student's learning	0.037
Private tutoring	0.035
Social support	0.029
Motivation level	0.025
4 English	0.022
2 Politics	0.019
2 Biology	0.019
3 English	0.015
2 Geography	0.015
2 Chinese	0.014
2 History	0.011
1 English	0.008
1 Politics	0.007

The initial prediction accuracy rate of Mathematics performance using ANNs was 69.9%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.30, continuing until the highest prediction accuracy is achieved.

Table 4.30: Factors Removed from Predicting Student NCEE Performance in Mathematics Subject Using ANNs

No.	Factors Removed	Prediction accuracy (%)
1	1 Politics	69.9
2	1 Politics, 1 English	69.9
3	1 Politics, 1 English, 2 History	69.9
4	1 Politics, 1 English, 2 History, 2 Chinese	71.7

Table 4.30 continued

5	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography	71.7
6	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English	74.6
7	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology	74.6
8	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics	74.6
9	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English	76.4
10	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level	76.4
11	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support	79.5
12	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring	79.5
13	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning	79.5
14	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English	79.5
15	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry	82.1
16	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics	82.1
17	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography	82.9
18	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores	82.9
19	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology	84.3

Table 4.30 continued

20	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores	85.0
21	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese	85.0
22	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics	85.0
23	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History	87.3
24	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics	87.3
25	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age	89.7
26	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age, 2 Chemistry	89.7
27	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age, 2 Chemistry, 3 Total scores	92.8
28	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age, 2 Chemistry, 3 Total scores, Teacher's education level	92.8

Table 4.30 continued

29	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History	92.8
30	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 1 Chemistry	92.8
31	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 1 Chemistry, 3 Chinese	94.0
32	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 1 Chemistry, 3 Chinese, 3 Politics	95.1
33	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 1 Chemistry, 3 Chinese, 3 Politics, Annual family income	95.1
34	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 1 Chemistry, 3 Chinese, 3 Politics, Annual family income, Teacher's qualification	95.1
35	1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, Motivation level, Social support, Private tutoring, Parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 Total scores, 3 Biology, 1 Total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, Age, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 1 Chemistry, 3 Chinese, 3 Politics, Annual family income, Teacher's qualification, Gender	81.9

According to Table 4.30, when factors, such as 1 Politics, 1 English, 2 History, 2 Chinese, 2 Geography, 3 English, 2 Biology, 2 Politics, 4 English, motivation level, social support, private tutoring, parental involvement in student's learning, 2 English, 3 Chemistry, 1 Mathematics, 3 Geography, 4 total scores, 3 Biology, 1 total scores, 1 Chinese, 1 Physics, 1 History, 2 Physics, age, 2 Chemistry, 3 total scores, teacher's education level, 3 History, 1 Chemistry, 3 Chinese, 3 Politics, annual family income, and teacher's qualification are removed, the prediction accuracy reaches 95.1%, while when Gender is removed, the prediction accuracy drops to 81.9%. From this, it can be known that factors such as test anxiety, mother's education level, father's education level, socioeconomic status, peer relationship, student's self-efficacy, teacher's self-efficacy, teaching method, 4 Mathematics, 3 Physics, 2 Mathematics, 3 Mathematics, and gender are suitable for predicting student NCEE performance in Mathematics subject using ANNs.

d) SVMs

Figure 4.8 shows the process of identifying the most important factors to predict student NCEE performance in the Mathematics subject using SVMs.

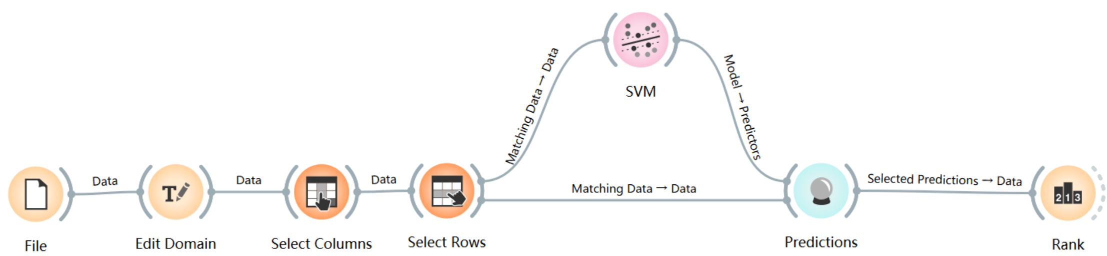

Figure 4.8: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the Mathematics Subject Using SVMs

Table 4.31 lists the weight of attributes for predicting student NCEE performance in the Mathematics subject using SVMs.

Table 4.31: Weight of the Attributes for Predicting Student NCEE Performance in Mathematics Subject Using SVMs

Attributes for Mathematics	Weight
Test anxiety	0.839
Father's education level	0.559
Mother's education level	0.526
Socioeconomic status	0.405
Student's self-efficacy	0.249
4 Mathematics	0.227
Teaching method	0.213
Teacher's self-efficacy	0.202
Peer relationship	0.189
Gender	0.128
3 Mathematics	0.124
1 Mathematics	0.124
2 Mathematics	0.113
4 Total scores	0.112
Annual family income	0.086
Teacher's qualification	0.086
3 Chemistry	0.078
4 English	0.077
3 Politics	0.075
1 Politics	0.074
Teacher's education level	0.069
2 Physics	0.067
3 Physics	0.060
1 Chinese	0.060
3 Chinese	0.058

Table 4.31 continued

1 History	0.057
2 Geography	0.054
Motivation level	0.053
1 Physics	0.051
3 Geography	0.050
2 Politics	0.045
2 Biology	0.045
2 Chemistry	0.044
1 Chemistry	0.040
2 Chinese	0.039
2 English	0.038
Age	0.037
1 English	0.037
3 History	0.036
4 Chinese	0.035
2 History	0.031
3 English	0.030
Parental involvement in student's learning	0.021
1 Total scores	0.012
Private tutoring	0.012
Social support	0.006

The initial prediction accuracy rate of Mathematics performance using SVMs was 72.1%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.32, continuing until the highest prediction accuracy is achieved.

Table 4.32: Factors Removed from Predicting Student NCEE Performance in Mathematics Subject Using SVMs

No.	Factors Removed	Prediction accuracy (%)
1	Social support	72.1
2	Social support, Private tutoring	72.1
3	Social support, Private tutoring, 1 Total scores	72.1
4	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning	72.1
5	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English	73.9
6	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History	73.9
7	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese	73.9
8	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History	74.4
9	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English	74.4
10	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age	76.3
11	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English	77.7
12	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese	77.7
13	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry	81.9
14	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry	81.9
15	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology	81.9

Table 4.32 continued

16	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics	85.2
17	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography	85.2
18	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics	85.2
19	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level	87.0
20	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography	87.7
21	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History	87.7
22	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese	90.7
23	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese	90.7
24	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics	90.7
25	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics, 2 Physics	90.7

Table 4.32 continued

26	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics, 2 Physics, Teacher's education level	92.0
27	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics, 2 Physics, Teacher's education level, 1 Politics	92.0
28	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics, 2 Physics, Teacher's education level, 1 Politics, 3 Politics	92.5
29	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics, 2 Physics, Teacher's education level, 1 Politics, 3 Politics, 4 English	92.5
30	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics, 2 Physics, Teacher's education level, 1 Politics, 3 Politics, 4 English, 3 Chemistry	92.5
31	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics, 2 Physics, Teacher's education level, 1 Politics, 3 Politics, 4 English, 3 Chemistry, Teacher's qualification	94.7
32	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics, 2 Physics, Teacher's education level, 1 Politics, 3 Politics, 4 English, 3 Chemistry, Teacher's qualification, Annual family income	94.7

Table 4.32 continued

33	Social support, Private tutoring, 1 Total scores, Parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, Age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, Motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics, 2 Physics, Teacher's education level, 1 Politics, 3 Politics, 4 English, 3 Chemistry, Teacher's qualification, Annual family income, 4 Total scores	80.6

According to Table 4.32, when factors, such as social support, private tutoring, 1 total scores, parental involvement in student's learning, 3 English, 2 History, 4 Chinese, 3 History, 1 English, age, 2 English, 2 Chinese, 1 Chemistry, 2 Chemistry, 2 Biology, 2 Politics, 3 Geography,1 Physics, motivation level, 2 Geography, 1 History, 3 Chinese, 1 Chinese, 3 Physics, 2 Physics, teacher's education level, 1 Politics, 3 Politics, 4 English, 3 Chemistry, teacher's qualification, annual family income are removed, the prediction accuracy reaches 94.7%, while when 4 total scores is removed, the prediction accuracy drops to 80.6%. From this, it can be known that factors such as test anxiety, father's education level, mother's education level, socioeconomic status, student's self-efficacy, 4 Mathematics, teaching method, teacher's self-efficacy, peer relationship, gender, 3 Mathematics, 1 Mathematics, 2 Mathematics and 4 total scores are suitable for predicting student NCEE performance in Mathematics subject using SVMs.

In order to obtain the potentially best predictor factors of NCEE performance in Mathematics, researcher temporarily take the top five affecting factors of each algorithm, and then use four algorithms to make a re-prediction and check the prediction accuracy. By aggregating the top five predictive factors of each of the four algorithms and then eliminating the duplicate parts, four new sets of predictive factors and weight can be obtained (see Table 4.33), and also the prediction accuracy (see Table 4.34).

Table 4.33: Four New Sets of Predictive Factors and Weight to Repredict Student NCEE Performance in Mathematics Subject using Four Algorithms

Factors using NB	Weight	Factors using DT	Weight	Factors using ANNs	Weight	Factors using SVMs	Weight
Test anxiety	0.793	Test anxiety	0.773	Test anxiety	0.742	Test anxiety	0.791
Father's education level	0.349	Father's education level	0.406	Father's education level	0.378	Father's education level	0.379
Mother's education level	0.335	Mother's education level	0.334	Mother's education level	0.364	Mother's education level	0.359
Socioeconomic status	0.246	Socioeconomic status	0.252	Socioeconomic status	0.280	Socioeconomic status	0.247
Peer relationship	0.107	Student's self-efficacy	0.160	Student's self-efficacy	0.166	Student's self-efficacy	0.191
Student's self-efficacy	0.085	Peer relationship	0.122	Peer relationship	0.133	Peer relationship	0.126

Table 4.34: The Prediction Accuracy of Repredicting Student NCEE Performance in Mathematics Subject using Four Algorithms

Algorithm	Prediction Accuracy (%)	Algorithm	Prediction Accuracy (%)	Algorithm	Prediction Accuracy (%)	Algorithm	Prediction Accuracy (%)
NB	94.4	DT	91.3	ANNs	92.6	SVMs	91.9

As fine-tuning did not improve the prediction accuracy, the initial prediction using NB, which achieved the highest accuracy (96.4%), was used to determine the predictor factors. Referring to Table 4.31, the predictors of NCEE performance in Mathematics subject are test anxiety (weight 0.839) is the most critical factor, followed by father's education level (weight 0.559), mother's education level (weight 0.526), socioeconomic status (weight 0.405), student's self-efficacy (weight 0.249), 4 Mathematics (weight 0.227), teaching method (weight 0.213), teacher's self-efficacy (weight 0.202), peer relationship (weight 0.189), gender (weight 0.128), 3 Mathematics (weight 0.124), 1 Mathematics (weight 0.124), 2 Mathematics (weight 0.113) and 4 total scores (weight 0.112). Some of these predictors are also predictors of the other three algorithms (see Table 4.35).

Table 4.35: Same Predictors using NB in Other Three Algorithms in Mathematics Subject

Factors using NB	Whether in DT	Whether in ANNs	Whether in SVMs
Test anxiety	Yes	Yes	Yes
Father's education level	Yes	Yes	Yes
Mother's education level	Yes	Yes	Yes
Socioeconomic status	Yes	Yes	Yes
Peer relationship	Yes	Yes	Yes
Student's self-efficacy	Yes	Yes	Yes
2 Mathematics	Yes	Yes	Yes
Teaching method	Yes	Yes	Yes
4 Mathematics	Yes	Yes	Yes
3 Physics	Yes	Yes	No
Teacher's self-efficacy	Yes	Yes	Yes
3 Chemistry	Yes	No	No
4 Total scores	No	No	Yes

4.5.3 The Most Important Factors to Predict Student NCEE Performance in English Subject

RQ1c: What are the most important factors to predict student NCEE performance in English subject based on NB, DT, ANNs, SVMs algorithms?

a) NB

Figure 4.9 shows the process of identifying the most important factors to predict student NCEE performance in the English subject using NB.

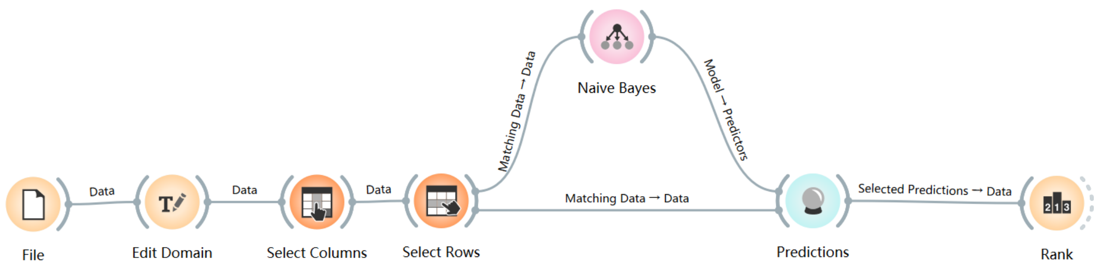

Figure 4.9: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the English Subject Using NB

Table 4.36 lists the weight of attributes for predicting student NCEE performance in the English subject using NB.

Table 4.36: Weight of the Attributes for Predicting Student NCEE Performance in English Subject Using NB

Attributes for English	Weight
Annual family income	0.605
Parental involvement in student's learning	0.411
4 English	0.264
3 English	0.160
2 English	0.159
Socioeconomic status	0.132
Teacher's qualification	0.116
Teacher's self-efficacy	0.115
Teaching method	0.110
4 Total scores	0.102
Motivation level	0.086
4 Chinese	0.085
Teacher's education level	0.078
2 History	0.077

Table 4.36 continued

2 Chinese	0.063
Father's education level	0.060
1 Physics	0.055
1 History	0.050
Social support	0.050
3 Geography	0.044
3 Politics	0.038
Mother's education level	0.035
2 Politics	0.035
2 Biology	0.035
Peer relationship	0.035
1 Total scores	0.035
3 Total scores	0.032
3 Chinese	0.030
1 Politics	0.029
Gender	0.029
2 Physics	0.023
1 English	0.023
1 Chemistry	0.018
2 Total scores	0.015
1 Mathematics	0.014
Student's self-efficacy	0.014
1 Chinese	0.012
2 Chemistry	0.009
4 Mathematics	0.009
Age	0.008
Test anxiety	0.008
3 Chemistry	0.003

The initial prediction accuracy rate of English performance using NB was 68.9%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.37, continuing until the highest prediction accuracy is achieved.

Table 4.37: Factors Removed from Predicting Student NCEE Performance in English Subject Using NB

No.	Factors Removed	Prediction accuracy (%)
1	3 Chemistry	68.9
2	3 Chemistry, Test anxiety	68.9
3	3 Chemistry, Test anxiety, Age	68.9
4	3 Chemistry, Test anxiety, Age, 4 Mathematics	68.9
5	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry	70.8
6	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese	70.8
7	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy	71.9
8	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics	73.0
9	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores	73.0
10	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry	74.8
11	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English	74.8
12	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics	74.8
13	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender	77.7

Table 4.37 continued

14	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics	77.7
15	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese	79.5
16	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores	79.5
17	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores	79.5
18	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship	82.6
19	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology	82.9
20	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics	82.9
21	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level	84.1
22	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics	84.1
23	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography	85.9
24	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography, Social support	85.9

Table 4.37 continued

25	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography, Social support, 1 History	85.9
26	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography, Social support, 1 History, 1 Physics	85.9
27	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography, Social support, 1 History, 1 Physics, Father's education level	87.7
28	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography, Social support, 1 History, 1 Physics, Father's education level, 2 Chinese	87.7
29	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography, Social support, 1 History, 1 Physics, Father's education level, 2 Chinese, 2 History	89.0
30	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography, Social support, 1 History, 1 Physics, Father's education level, 2 Chinese, 2 History, Teacher's education level	90.7
31	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography, Social support, 1 History, 1 Physics, Father's education level, 2 Chinese, 2 History, Teacher's education level, 4 Chinese	90.7
32	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography, Social support, 1 History, 1 Physics, Father's education level, 2 Chinese, 2 History, Teacher's education level, 4 Chinese, Motivation level	90.7

Table 4.37 continued

33	3 Chemistry, Test anxiety, Age, 4 Mathematics, 2 Chemistry, 1 Chinese, Student's self-efficacy, 1 Mathematics, 2 Total scores, 1 Chemistry, 1 English, 2 Physics, Gender, 1 Politics, 3 Chinese, 3 Total scores, 1 Total scores, Peer relationship, 2 Biology, 2 Politics, Mother's education level, 3 Politics, 3 Geography, Social support, 1 History, 1 Physics, Father's education level, 2 Chinese, 2 History, Teacher's education level, 4 Chinese, Motivation level, 4 Total scores	83.3

According to Table 4.37, when factors, such as 3 Chemistry, test anxiety, age, 4 Mathematics, 2 Chemistry, 1 Chinese, student's self-efficacy, 1 Mathematics, 2 total scores, 1 Chemistry, 1 English, 2 Physics, gender, 1 Politics, 3 Chinese, 3 total scores, 1 total scores, peer relationship, 2 Biology, 2 Politics, mother's education level, 3 Politics, 3 Geography, social support, 1 History, 1 Physics, father's education level, 2 Chinese, 2 History, teacher's education level, 4 Chinese, and motivation level are removed, the prediction accuracy reaches 90.7%, while when 4 total scores is removed, the prediction accuracy drops to 83.3%. From this, it can be known that factors such as annual family income, parental involvement in student's learning, 4 English, 3 English, 2 English, socioeconomic status, teacher's qualification, teacher's self-efficacy, teaching method and 4 total scores are suitable for predicting student NCEE performance in English subject using NB.

b) DT

Figure 4.10 shows the process of identifying the most important factors to predict student NCEE performance in the English subject using DT.

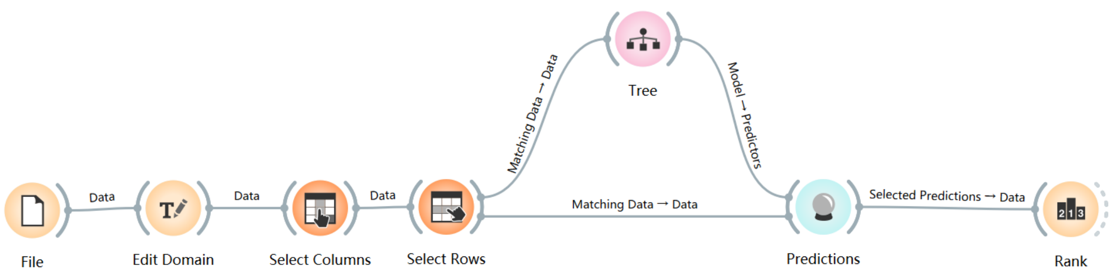

Figure 4.10: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the English Subject Using DT

Table 4.38 lists the weight of attributes for predicting student NCEE performance in the English subject using DT.

Table 4.38: Weight of the Attributes for Predicting Student NCEE Performance in English Subject Using DT

Attributes for English	Weight
Annual family income	0.680
Parental involvement in student's learning	0.493
4 English	0.390
3 English	0.230
2 English	0.223
Socioeconomic status	0.194
Teaching method	0.178
Teacher's self-efficacy	0.161
1 English	0.149
4 Total scores	0.142
Motivation level	0.136
Teacher's qualification	0.133
4 Chinese	0.132
Gender	0.129
3 Geography	0.103
3 Chinese	0.080
2 Geography	0.078
3 Politics	0.076
Teacher's education level	0.070

Table 4.38 continued

Age	0.070
2 History	0.062
1 Chemistry	0.062
3 Biology	0.060
Social support	0.059
1 History	0.053
Mother's education level	0.050
Father's education level	0.048
1 Physics	0.047
Test anxiety	0.042
3 History	0.041
2 Chinese	0.040
3 Chemistry	0.037
1 Politics	0.034
1 Mathematics	0.034
Peer relationship	0.032
4 Mathematics	0.031
1 Chinese	0.030
2 Physics	0.027
3 Physics	0.025
3 Total scores	0.022
Student's self-efficacy	0.019
2 Politics	0.017
2 Biology	0.017
2 Chemistry	0.013
3 Mathematics	0.008
1 Total scores	0.005
2 Total scores	0.005

Table 4.38 continued

Private tutoring	0.002

The initial prediction accuracy rate of English performance using DT was 70.0%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.39, continuing until the highest prediction accuracy is achieved.

Table 4.39: Factors Removed from Predicting Student NCEE Performance in English Subject Using DT

No.	Factors Removed	Prediction accuracy (%)
1	Private tutoring	70.0
2	Private tutoring, 2 Total scores	70.0
3	Private tutoring, 2 Total scores, 1 Total scores	70.0
4	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics	70.0
5	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry	72.9
6	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology	72.9
7	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics	72.9
8	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student’s self-efficacy	73.8
9	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student’s self-efficacy, 3 Total scores	73.8
10	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student’s self-efficacy, 3 Total scores, 3 Physics	73.8
11	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student’s self-efficacy, 3 Total scores, 3 Physics, 2 Physics	75.8

Table 4.39 continued

12	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese	76.6
13	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics	76.6
14	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship	79.8
15	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics	79.8
16	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics	79.8
17	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry	80.4
18	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese	80.4
19	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History	82.1
20	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety	83.8
21	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics	83.8
22	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level	83.8

Table 4.39 continued

23	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level	85.1
24	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History	85.1
25	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History, Social support	85.6
26	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History, Social support, 3 Biology	85.6
27	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History, Social support, 3 Biology, 1 Chemistry	85.6
28	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History, Social support, 3 Biology, 1 Chemistry, 2 History	85.6
29	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History, Social support, 3 Biology, 1 Chemistry, 2 History, Age	87.1
30	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History, Social support, 3 Biology, 1 Chemistry, 2 History, Age, Teacher's education level	87.1

Table 4.39 continued

31	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History, Social support, 3 Biology, 1 Chemistry, 2 History, Age, Teacher's education level, 3 Politics	88.4
32	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History, Social support, 3 Biology, 1 Chemistry, 2 History, Age, Teacher's education level, 3 Politics, 2 Geography	88.4
33	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History, Social support, 3 Biology, 1 Chemistry, 2 History, Age, Teacher's education level, 3 Politics, 2 Geography, 3 Chinese	88.4
34	Private tutoring, 2 Total scores, 1 Total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, Student's self-efficacy, 3 Total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, Peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, Test anxiety, 1 Physics, Father's education level, Mother's education level, 1 History, Social support, 3 Biology, 1 Chemistry, 2 History, Age, Teacher's education level, 3 Politics, 2 Geography, 3 Chinese, 3 Geography	76.8

According to Table 4.39, when factors, such as private tutoring, 2 total scores, 1 total scores, 3 Mathematics, 2 Chemistry, 2 Biology, 2 Politics, student's self-efficacy, 3 total scores, 3 Physics, 2 Physics, 1 Chinese, 4 Mathematics, peer relationship, 1 Mathematics, 1 Politics, 3 Chemistry, 2 Chinese, 3 History, test anxiety, 1 Physics, father's education level, mother's education level, 1 History, social support, 3 Biology, 1 Chemistry, 2 History, age, teacher's education level, 3 Politics, 2 Geography, 3 Chinese are removed, the prediction accuracy reaches 88.4%, while when 3 Geography is removed, the prediction accuracy drops to 76.8%. From this, it can be known that factors such as annual family income, parental involvement in student's learning, 4 English, 3 English, 2 English, socioeconomic status, teaching method, teacher's self-efficacy, 1 English, 4 total scores, motivation level, teacher's qualification, 4 Chinese, gender, and 3 Geography are suitable for predicting student NCEE performance in English subject using DT.

c) ANNs

Figure 4.11 shows the process of identifying the most important factors to predict student NCEE performance in the English subject using ANNs.

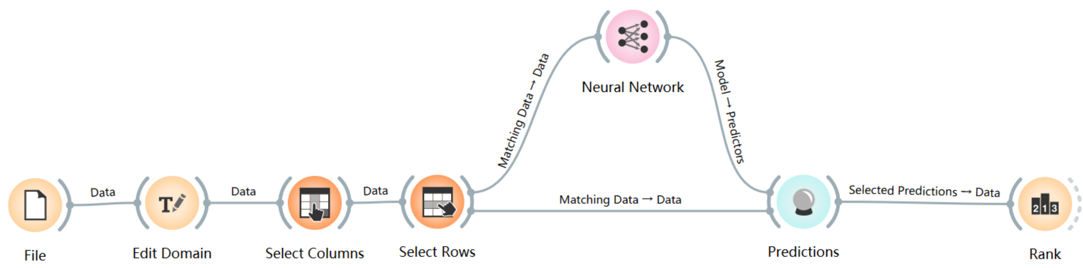

Figure 4.11: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the English Subject Using ANNs

Table 4.40 lists the weight of attributes for predicting student NCEE performance in the English subject using ANNs.

Table 4.40: Weight of the Attributes for Predicting Student NCEE Performance in English Subject Using ANNs

Attributes for English	Weight
Parental involvement in student's learning	0.551
Annual parental family income	0.520
4 English	0.308
3 English	0.219
2 English	0.198

Table 4.40 continued

1 English	0.144
Teacher's qualification	0.140
4 Total scores	0.138
Socioeconomic status	0.105
Motivation level	0.101
Teaching method	0.101
3 Politics	0.094
Teacher's self-efficacy	0.079
3 Chinese	0.076
2 Chinese	0.072
3 Geography	0.060
2 Politics	0.060
2 Biology	0.060
Gender	0.059
Father's education level	0.049
1 History	0.048
2 Geography	0.048
Teacher's education level	0.046
Age	0.042
Social support	0.038
3 Biology	0.037
Peer relationship	0.034
2 Total scores	0.034
4 Mathematics	0.033
3 Chemistry	0.032
4 Chinese	0.031
1 Politics	0.026
Test anxiety	0.023

Table 4.40 continued

2 History	0.020
1 Mathematics	0.019
2 Physics	0.019
1 Chemistry	0.017
3 Total scores	0.016
1 Chinese	0.011
2 Chemistry	0.003

The initial prediction accuracy rate of English performance using ANNs was 70.9%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.41, continuing until the highest prediction accuracy is achieved.

Table 4.41: Factors Removed from Predicting Student NCEE Performance in English Subject Using ANNs

No.	Factors Removed	Prediction accuracy (%)
1	2 Chemistry	70.9
2	2 Chemistry, 1 Chinese	70.9
3	2 Chemistry, 1 Chinese, 3 Total scores	70.9
4	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry	73.8
5	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics	73.8
6	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics	75.9
7	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History	75.9
8	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety	78.4
9	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics	78.4

Table 4.41 continued

10	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese	78.4
11	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry	79.3
12	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics	80.6
13	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores	80.6
14	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship	80.6
15	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology	81.7
16	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support	82.9
17	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age	82.9
18	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level	82.9
19	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography	82.9
20	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History	85.0
21	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History, Father's education level	85.0

Table 4.41 continued

22	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History, Father's education level, Gender	86.6
23	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History, Father's education level, Gender, 2 Biology	86.6
24	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History, Father's education level, Gender, 2 Biology, 2 Politics	86.6
25	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History, Father's education level, Gender, 2 Biology, 2 Politics, 3 Geography	88.1
26	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History, Father's education level, Gender, 2 Biology, 2 Politics, 3 Geography, 2 Chinese	88.1
27	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History, Father's education level, Gender, 2 Biology, 2 Politics, 3 Geography, 2 Chinese, 3 Chinese	88.1
28	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History, Father's education level, Gender, 2 Biology, 2 Politics, 3 Geography, 2 Chinese, 3 Chinese, Teacher's self-efficacy	89.3
29	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History, Father's education level, Gender, 2 Biology, 2 Politics, 3 Geography, 2 Chinese, 3 Chinese, Teacher's self-efficacy, 3 Politics	89.3

Table 4.41 continued

30	2 Chemistry, 1 Chinese, 3 Total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, Test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 Total scores, Peer relationship, 3 Biology, Social support, Age, Teacher's education level, 2 Geography, 1 History, Father's education level, Gender, 2 Biology, 2 Politics, 3 Geography, 2 Chinese, 3 Chinese, Teacher's self-efficacy, 3 Politics, Teaching method	71.7

According to Table 4.41, when factors, such as 2 Chemistry, 1 Chinese, 3 total scores, 1 Chemistry, 2 Physics, 1 Mathematics, 2 History, test anxiety, 1 Politics, 4 Chinese, 3 Chemistry, 4 Mathematics, 2 total scores, peer relationship, 3 Biology, Social support, age, teacher's education level, 2 Geography, 1 History, father's education level, gender, 2 Biology, 2 Politics, 3 Geography, 2 Chinese, 3 Chinese, teacher's self-efficacy, and 3 Politics are removed, the prediction accuracy reaches 89.3%, while when teaching method is removed, the prediction accuracy drops to 71.7%. From this, it can be known that factors such as parental involvement in student's learning, annual family income, 4 English, 3 English, 2 English, 1 English, teacher's qualification, 4 total scores, socioeconomic status, motivation level, and teaching method are suitable for predicting student NCEE performance in English subject using ANNs.

d) SVMs

Figure 4.12 shows the process of identifying the most important factors to predict student NCEE performance in the English subject using SVMs.

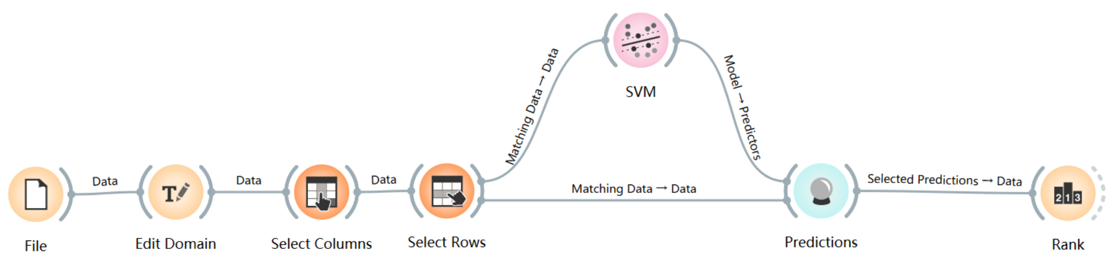

Figure 4.12: Process of Identifying the Most Important Factors to Predict Student NCEE Performance in the English Subject Using SVMs

Table 4.42 lists the weight of attributes for predicting student NCEE performance in the English subject using SVMs.

Table 4.42: Weight of the Attributes for Predicting Student NCEE Performance in English Subject Using SVMs

Attributes for English	Weight
Annual family income	0.635
Parental involvement in student's learning	0.366
4 English	0.350
Socioeconomic status	0.198
Teacher's qualification	0.186
Teaching method	0.170
Teacher's self-efficacy	0.149
2 English	0.146
3 English	0.138
3 Politics	0.131
Motivation level	0.125
2 Chinese	0.114
1 English	0.113
4 Total scores	0.110
Mother's education level	0.109
1 Mathematics	0.105
4 Chinese	0.105
Age	0.101
3 Geography	0.095

Table 4.42 continued

Father's education level	0.094
Gender	0.091
1 Politics	0.087
2 Geography	0.080
Test anxiety	0.080
1 Chemistry	0.059
1 Physics	0.054
1 History	0.054
Peer relationship	0.050
2 Politics	0.049
2 Biology	0.049
3 Biology	0.048
4 Mathematics	0.043
2 History	0.042
Social support	0.039
3 Chemistry	0.038
Student's self-efficacy	0.037
3 Chinese	0.028
3 History	0.027
Teacher's education level	0.018
3 Total scores	0.016
2 Chemistry	0.012
1 Chinese	0.012
3 Mathematics	0.010
2 Mathematics	0.010
2 Physics	0.005
2 Total scores	0.001

The initial prediction accuracy rate of English performance using SVMs was 73.4%. In order to identify the most significant affecting factors, the ones with low weights need to be removed. The elimination process is shown in Table 4.43, continuing until the highest prediction accuracy is achieved.

Table 4.43: Factors Removed from Predicting Student NCEE Performance in English Subject Using SVMs

No.	Factors Removed	Prediction accuracy (%)
1	2 Total scores	73.4
2	2 Total scores, 2 Physics	73.4
3	2 Total scores, 2 Physics, 2 Mathematics	73.4
4	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics	76.8
5	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese	76.8
6	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry	76.8
7	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores	79.7
8	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level	79.7
9	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History	80.6
10	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese	81.3
11	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy	81.3
12	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry	81.3
13	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support	84.0

Table 4.43 continued

14	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History	84.0
15	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics	85.4
16	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology	85.4
17	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology	85.4
18	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics	88.8
19	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship	88.8
20	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship, 1 History	89.1
21	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship, 1 History, 1 Physics	89.1
22	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship, 1 History, 1 Physics, 1 Chemistry	89.1
23	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship, 1 History, 1 Physics, 1 Chemistry, Test anxiety	89.1

Table 4.43 continued

24	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship, 1 History, 1 Physics, 1 Chemistry, Test anxiety, 2 Geography	89.1
25	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship, 1 History, 1 Physics, 1 Chemistry, Test anxiety, 2 Geography, 1 Politics	90.0
26	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship, 1 History, 1 Physics, 1 Chemistry, Test anxiety, 2 Geography, 1 Politics, Gender	90.2
27	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship, 1 History, 1 Physics, 1 Chemistry, Test anxiety, 2 Geography, 1 Politics, Gender, Father's education level	90.2
28	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship, 1 History, 1 Physics, 1 Chemistry, Test anxiety, 2 Geography, 1 Politics, Gender, Father's education level, 3 Geography	90.2
29	2 Total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 Total scores, Teacher's education level, 3 History, 3 Chinese, Student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, Peer relationship, 1 History, 1 Physics, 1 Chemistry, Test anxiety, 2 Geography, 1 Politics, Gender, Father's education level, 3 Geography, Age	81.7

According to Table 4.43, when factors, such as 2 total scores, 2 Physics, 2 Mathematics, 3 Mathematics, 1 Chinese, 2 Chemistry, 3 total scores, teacher's education level, 3 History, 3 Chinese, student's self-efficacy, 3 Chemistry, Social support, 2 History, 4 Mathematics, 3 Biology, 2 Biology, 2 Politics, peer relationship, 1 History, 1 Physics, 1 Chemistry, test anxiety, 2 Geography, 1 Politics, Gender, father's education level, 3 Geography are removed, the prediction accuracy reaches 90.2%, while when age is removed, the prediction accuracy drops to 79.6%. From this, it can be known that factors such as annual family income, parental involvement in student's learning, 4 English, socioeconomic status, teacher's qualification, teaching method, teacher's self-efficacy, 2 English, 3 English, 3 Politics, motivation level, 2 Chinese, 1 English, 4 total scores, mother's education level, 1 Mathematics, 4 Chinese, and age are suitable for predicting student NCEE performance in English subject using SVMs.

In order to obtain the potentially best predictor factors of NCEE performance in English, researcher will temporarily take the top five affecting factors of each algorithm and then use four algorithms to make a re-prediction and check the prediction accuracy. By aggregating the top five predictive factors of each of the four algorithms and then eliminating the duplicate parts, four new sets of predictive factors and weight can be obtained (see Table 4.44), and also the prediction accuracy (see Table 4.45).

Table 4.44: Four New Sets of Predictive Factors and Weight to Repredict Student NCEE Performance in English Subject using Four Algorithms

Factors using NB	Weight	Factors using DT	Weight	Factors using ANNs	Weight	Factors using SVMs	Weight
Annual family income	0.476	Parental involvement in student's learning	0.522	Annual family income	0.514	Parental involvement in student's learning	0.426
Parental involvement in student's learning	0.392	Annual family income	0.417	Parental involvement in student's learning	0.419	Annual family income	0.417
4 English	0.246	4 English	0.255	4 English	0.182	4 English	0.290
Teacher's qualification	0.143	Teacher's qualification	0.226	Socioeconomic status	0.166	2 English	0.153
Socioeconomic status	0.104	Socioeconomic status	0.114	Teacher's qualification	0.158	3 English	0.147
2 English	0.098	3 English	0.111	3 English	0.148	Socioeconomic status	0.146

Table 4.44 continued

3 English	0.080	2 English	0.014	2 English	0.074	Teacher's qualification	0.110

Table 4.45: The Prediction Accuracy of Repredicting Student NCEE Performance in English Subject using Four Algorithms

Algorithm	Prediction Accuracy (%)	Algorithm	Prediction Accuracy (%)	Algorithm	Prediction Accuracy (%)	Algorithm	Prediction Accuracy (%)
NB	89.8	DT	87.6	ANNs	88.7	SVMs	89.5

As fine-tuning did not improve the prediction accuracy, the initial prediction using NB, which achieved the highest accuracy (90.7%), was used to determine the predictor factors. Referring to Table 4.36, the predictors of NCEE performance in English subject are annual family income (weight 0.605) is the most critical factor, followed by f parental involvement in student's learning (weight 0.411), 4 English (weight 0.264), 3 English (weight 0.160), 2 English (weight 0.159), socioeconomic status(weight 0.132), teacher's qualification (weight 0.116), teacher's self-efficacy (weight 0.115), teaching method (weight 0.110) and 4 total scores (weight 0.102). Some of these predictors are also predictors of the other three algorithms (see Table 4.36).

Table 4.46: Same Predictors using NB in Other Three Algorithms in English Subject

Factors using NB	Whether in DT	Whether in ANNs	Whether in SVMs
Annual family income	Yes	Yes	Yes
Parental involvement in student's learning	Yes	Yes	Yes
4 English	Yes	Yes	Yes
3 English	Yes	Yes	Yes
2 English	Yes	Yes	Yes

Table 4.46 continued

Socioeconomic status	Yes	Yes	Yes
Teacher's qualification	Yes	Yes	Yes
Teacher's self-efficacy	Yes	No	Yes
Teaching method	Yes	Yes	Yes
4 Total scores	Yes	Yes	Yes

4.6 Models Performance Evaluation

Based on the results of Section 4.5, it can be concluded that before and after fine-tuning the model, the accuracy of the four algorithms (NB, DT, ANNs, SVMs) in predicting the performance of Chinese, Mathematics and English changed, as well as the accuracy after the re-prediction, as shown in Table 4.47.

Table 4.47: Accuracy of Model Before and After Fine-Tuning and Re-Prediction

	Before and After Fine-Tuning	Piediction of Chinese (%)	Re-Prediction of Chinese (%)	Piediction of Mathematics (%)	Re-Prediction of Mathematics (%)	Piediction of Mathematics (%)	Re-Prediction of Mathematics (%)
NB	Before	68.1	92.1	70.2	94.4	68.9	89.8
	After	95.1		94.6		90.7	
DT	Before	72.5	90.9	71	91.3	70.0	87.6
	After	92.9		92.9		88.4	
ANNs	Before	78.4	91.7	69.9	92.6	70.9	88.7
	After	93.8		95.1		89.3	
SVMs	Before	69.8	89.7	72.1	91.9	73.4	89.5
	After	92.9		94.7		90.2	

As can be seen from Table 4.47, the accuracy of the re-prediction is generally lower than that of the model after fine-tuning. Therefore, the prediction still proceeds according to the fine-tuned model. At this time, the most important predictive factors, according to the results of Section 4.5, are summarized as shown in Table 4.48.

Table 4.48: Main Predictive Factors after Model Fine-Tuning

Subject	Algorithm	Main Predictive Factors after Model Fine-Tuning
Chinese	NB	Motivation level, teaching method, gender, 4 Chinese, teacher's self-efficacy, 4 English, 3 Chinese, 2 Politics, 2 Biology, and teacher's qualification
	DT	Motivation level, teaching method, teacher's qualification, 4 English, annual family income, parental involvement in student's learning, gender, teacher's self-efficacy, 3 Chinese, socioeconomic status, teacher's education level, 3 Physics, and 4 Chinese
	ANNs	Motivation level, 3 Chinese, teacher's qualification, teacher's self-efficacy, parental involvement in student's learning, 2 Chinese, 4 Chinese, teacher's education level, teaching method, gender, mother's education level, 1 Mathematics and peer relationship
	SVMs	Motivation level, teaching method, teacher's self-efficacy, teacher's qualification, age, 4 English, 2 Chinese, 3 Chinese, parental involvement in student's learning, 1 total scores, teacher's education level, and 1 English
Mathematics	NB	Test anxiety, father's education level, mother's education level, socioeconomic status, peer relationship, student's self-efficacy, 2 Mathematics, teaching method, 4 Mathematics, 3 Physics, teacher's self-efficacy, 3 Chemistry, and total scores
	DT	Test anxiety, father's education level, mother's education level, socioeconomic status, student's self-efficacy, peer relationship, 4 Mathematics, 2 Mathematics, teacher's self-efficacy, teaching method, 3 Mathematics, 3 Physics, 1 Mathematics, gender, teacher's qualification, teacher's education level, and 3 Chemistry
	ANNs	Test anxiety, mother's education level, father's education level, socioeconomic status, peer relationship, student's self-efficacy, teacher's self-efficacy, teaching method, 3 Physics, 2 Mathematics, 3 Mathematics, and gender
	SVMs	Test anxiety, father's education level, mother's education level, socioeconomic status, student's self-efficacy, 4 Mathematics, teaching method, teacher's self-efficacy, peer relationship, gender, 3 Mathematics, 1 Mathematics, 2 Mathematics and 4 total scores
English	NB	Annual family income, parental involvement in student's learning, 4 English, 3 English, 2 English, socioeconomic status, teacher's qualification, teacher's self-efficacy, teaching method and 4 total scores
	DT	Annual family income, parental involvement in student's learning, 4 English, 3 English, 2 English, socioeconomic status, teaching method, teacher's self-efficacy, 1 English, 4 total scores, motivation level, teacher's qualification, 4 Chinese, gender, and 3 Geography
	ANNs	Parental involvement in student's learning, annual family income, 4 English, 3 English, 2 English, 1 English, teacher's qualification, 4 total scores, socioeconomic status, motivation level, and teaching method
	SVMs	Annual family income, parental involvement in student's learning, 4 English, socioeconomic status, teacher's qualification, teaching method, teacher's self-efficacy, 2 English, 3 English, 3 Politics, motivation level, 2 Chinese, 1 English, 4 total scores, mother's education level, 1 Mathematics, 4 Chinese, and age

The confusion matrix shows how well a model predicts student performance by comparing true and predicted results (see Table 4.49).

Table 4.49: Confusion Matrix of Predicting Student Performance

		Predicted class	
		Positive	Negative
Actual class	Positive	TP	FN
	Negative	FP	TN

In Table 4.49, TP is when the model correctly identifies a positive case. FN happens if a positive case is wrongly predicted as negative. FP occurs when the model falsely predicts a positive case. TN refers to the model accurately identifying a negative case. Based on the confusion matrix, the measures for predicting student performance shown in Table 4.50 can be obtained.

Table 4.50: Measures in Predicting Student Performance Based on the Confusion Matrix

Criterion	**Formula**
AUC	AUC (Area Under the Curve) refers to the area under the ROC (Receiver Operating Characteristic) curve.
Accuracy	(TP+TN)/(TP+TN+FP+FN)
F1-Score	2TP/(2TP+FP+FN)
Precision	TP/(TP+FP)
Recall	TP/(TP+FN)
MCC	$(TP \cdot TN - FP \cdot FN)/\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}$

4.6.1 The Highest Accuracy Predictive Model in Predicting Student NCEE Chinese Performance

RQ2a: Which predictive model has the highest accuracy in predicting Chinese performance using NB, DT, ANNs, SVMs algorithms?

a) NB

The confusion matrix in Table 4.51 shows the results of NB algorithm in predicting Chinese performance. The model correctly classified 214 out of 225 students, achieving an accuracy of 95.1%. Among low-performing students, 132 were correctly identified, while 6 were misclassified as high. For moderate performers, 49 were accurately predicted. In the high group, 33 students were correctly identified, with 3 misclassified as low. Misclassifications mainly occurred between distant categories.

Table 4.51: Confusion Matrix in Predicting Chinese Performance Using NB

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
Actual class	Ture Low	**132**	0	6	138
	Ture Moderate	1	**49**	1	51
	Ture High	3	0	**33**	36
	Total	136	49	40	225

Table 4.52 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting Chinese performance using NB.

Table 4.52: Measures in Predicting Chinese Performance Using NB

Criterion	Value (%)
AUC	98.7
Accuracy	95.1
F1-Score	95.2
Precision	95.4
Recall	95.1
MCC	91.2

b) DT

The confusion matrix in Table 4.53 shows the results of the DT algorithm in predicting Chinese performance. The model correctly predicted 209 out of 225 cases, with an overall accuracy of 92.9%. Among low performers, 126 were correctly identified; 12 were misclassified as high. All moderate students were predicted correctly except for 2. For high performers, 34 were accurately classified, while 2 were labelled as low.

Table 4.53: Confusion Matrix in Predicting Chinese Performance Using DT

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
Actual class	Ture Low	**126**	0	12	138
	Ture Moderate	0	**49**	2	51
	Ture High	2	0	**34**	36
	Total	128	49	48	225

Table 4.54 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting Chinese performance using DT.

Table 4.54: Measures in Predicting Chinese Performance Using DT

Criterion	**Value (%)**
AUC	98.2
Accuracy	92.9
F1-Score	93.3
Precision	94.4
Recall	92.9
MCC	87.9

c) ANNs

The confusion matrix in Table 4.55 illustrates the results of ANNs algorithm in predicting Chinese performance. The model correctly predicted 211 out of 225 cases, yielding an accuracy of 93.8%. Among low-performing students, 132 were correctly identified; 6 were misclassified as high. For moderate performers, 48 were predicted accurately, and for high performers, 31 were correctly classified, with 4 misclassified as low.

Table 4.55: Confusion Matrix in Predicting Chinese Performance Using ANNs

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
	Ture Low	**132**	0	6	138
Actual class	Ture Moderate	2	**48**	1	51
	Ture High	4	1	**31**	36
	Total	138	49	38	225

Table 4.56 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting Chinese performance using ANNs.

Table 4.56: Measures in Predicting Chinese Performance Using ANNs

Criterion	**Value (%)**
AUC	98.5
Accuracy	93.8
F1-Score	93.8
Precision	93.9
Recall	93.8
MCC	88.6

d) SVMs

The confusion matrix in Table 4.57 illustrates the results of SVMs algorithm in predicting Chinese performance. The model correctly classified 209 out of 225 cases, with an accuracy of 92.9%. Among low performers, 130 were predicted correctly; 8 were misclassified as high. For moderate students, 49 were accurate. Among high performers, 30 were correctly identified, with 5 labelled as low and 1 as moderate.

Table 4.57: Confusion Matrix in Predicting Chinese Performance Using SVMs

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
Actual class	Ture Low	**130**	0	8	138
	Ture Moderate	1	**49**	1	51
	Ture High	5	1	**30**	36
	Total	136	50	39	225

Table 4.58 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting Chinese performance using SVMs.

Table 4.58: Measures in Predicting Chinese Performance Using SVMs

Criterion	**Value (%)**
AUC	97.9
Accuracy	92.9
F1-Score	93.0
Precision	93.1
Recall	92.9
MCC	87.1

Table 4.59 summarizes the measures of four DM algorithms in predicting students' Chinese

performance.

Table 4.59: The Measures of Four DM Algorithms in Predicting Students' Chinese Performance

DM algorithms	Prediction Measures (%)					
	ACC	Accuracy	F1-Score	Precision	Recall	MCC
NB	98.7	95.1	95.2	95.4	95.1	91.2
DT	98.2	92.9	93.3	94.4	92.9	87.9
ANNs	98.5	93.8	93.8	93.9	93.8	88.6
SVMs	97.9	92.9	93.0	93.1	92.9	87.1

Table 4.59 shows that the highest accuracy was achieved by NB, at 95.1%. ANNs follows closely with 93.8%, while DT and SVMs have the lowest accuracy at 92.9%. This suggests that NB is better suited for predicting Chinese performance compared to the other models.

4.6.2 The Highest Accuracy Predictive Model in Predicting Student NCEE Mathematics Performance

RQ2b: Which predictive model has the highest accuracy in predicting Mathematics performance using NB, DT, ANNs, SVMs algorithms?

a) NB

The confusion matrix in table 4.60 shows the results of NB algorithm in predicting Mathematics performance. The model correctly predicted 217 out of 225 Mathematics performance cases, with an overall accuracy of 96.4%. Among low-performing students, 134 were accurately identified, with 2 misclassified as high. All 41 moderate students were correctly predicted except for 2 labelled as low. For high performers, 42 were correctly classified, while 3 were misclassified as low and 1 as moderate.

Table 4.60: Confusion Matrix in Predicting Mathematics Performance Using NB

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
Actual class	Ture Low	**134**	0	2	136
	Ture Moderate	2	**41**	0	43
	Ture High	3	1	**42**	46
	Total	139	42	44	225

Table 4.61 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting Mathematics performance using NB.

Table 4.61: Measures in Predicting Mathematics Performance Using NB

Criterion	**Value (%)**
AUC	99.3
Accuracy	96.4
F1-Score	96.4
Precision	96.4
Recall	96.4
MCC	93.6

b) DT

The confusion matrix in table 4.62 shows the results of DT algorithm in predicting Mathematics performance. The model achieved an accuracy of 92.9%, correctly predicting 209 out of 225 students' Mathematics performance. Among low-performing students, 134 were correctly classified, with 1 misclassified as moderate and 1 as high. For moderate students, 40 were accurately predicted, while 3 were misclassified as low. In the high-performing group, 35 were correctly identified, with 9 predicted as low and 2 as moderate. Misclassifications mainly occurred between low and high categories.

Table 4.62: Confusion Matrix in Predicting Mathematics Performance Using DT

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
Actual class	Ture Low	**134**	1	1	136
	Ture Moderate	3	**40**	0	43
	Ture High	9	2	**35**	46
	Total	146	43	36	225

Table 4.63 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting Mathematics performance using DT.

Table 4.63: Measures in Predicting Mathematics Performance Using DT

Criterion	Value (%)
AUC	97.3
Accuracy	92.9
F1-Score	92.7
Precision	93.1
Recall	92.9
MCC	87.2

c) ANNs

The confusion matrix in table 4.64 shows the results of ANNs algorithm in predicting Mathematics performance. The model correctly predicted 214 out of 225 cases, with an accuracy of 95.1%. Among low-performing students, 135 were correctly classified, with 1 misclassified as high.

For moderate performers, 40 were accurate, and 3 were misclassified as low. In the high-performing group, 39 were correctly identified, while 6 were misclassified as low and 1 as moderate.

Table 4.64: Confusion Matrix in Predicting Mathematics Performance Using ANNs

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
Actual class	Ture Low	**135**	0	1	136
	Ture Moderate	3	**40**	0	43
	Ture High	6	1	**39**	46
	Total	144	41	40	225

Table 4.65 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting Mathematics performance using ANNs.

Table 4.65: Measures in Predicting Mathematics Performance Using ANNs

Criterion	Value (%)
AUC	99.2
Accuracy	95.1
F1-Score	95.0
Precision	95.2
Recall	95.1
MCC	91.2

d) SVMs

The confusion matrix in table 4.66 shows the results of SVMs algorithm in predicting Mathematics performance. The model correctly classified 213 out of 225 students, yielding an accuracy of 94.7%. Among low-performing students, 135 were accurately identified, with 1 misclassified as high. For moderate performers, 39 were correct, while 3 were misclassified as low and 1 as high. In the high-performing group, 39 were correctly predicted, with 5 misclassified as low and 2 as moderate.

Table 4.66: Confusion Matrix in Predicting Mathematics Performance Using SVMs

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
Actual class	Ture Low	**135**	0	1	136
	Ture Moderate	3	**39**	1	43
	Ture High	5	2	**39**	46
	Total	143	41	41	225

Table 4.67 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting Mathematics performance using SVMs.

Table 4.67: Measures in Predicting Chinese Performance Using SVMs

Criterion	Value (%)
AUC	99.3
Accuracy	94.7
F1-Score	94.6
Precision	94.7
Recall	94.7
MCC	90.3

Table 4.68 summarizes the measures of four DM algorithms in predicting students' Mathematics performance.

Table 4.68: The Measures of Four DM Algorithms in Predicting Students' Mathematics Performance

DM algorithms	Prediction Measures (%)					
	ACC	Accuracy	F1-Score	Precision	Recall	MCC
NB	99.3	96.4	96.4	96.4	96.4	93.6
DT	97.3	92.9	92.7	93.1	92.9	87.2
ANNs	99.2	95.1	95.0	95.2	95.1	91.2
SVMs	99.3	94.7	94.6	94.7	94.7	90.3

Table 4.68 indicates that NB achieved the highest accuracy at 96.4%, followed by ANNs (95.1%) and SVMs (94.7%). DT showed the lowest accuracy at 92.9%. This comparison highlights that NB performed best in predicting Mathematics performance.

4.6.3 The Highest Accuracy Predictive Model in Predicting Student NCEE English Performance

RQ2c: Which predictive model has the highest accuracy in predicting English performance using NB, DT, ANNs, SVMs algorithms?

a) NB

The confusion matrix in table 4.69 shows the results of NB algorithm in predicting English performance. The model achieved an accuracy of 90.7%, correctly classifying 204 out of 225 students. Among low performers, 135 were predicted correctly, with 9 misclassified. For moderate students, 42 were accurate, while 4 were misclassified. In the high group, 27 were correctly identified, and 8 were predicted as low or moderate.

Table 4.69: Confusion Matrix in Predicting English Performance Using NB

	Predicted class			
	Predicted Low	Predicted Moderate	Predicted High	Total

Table 4.69 continued

Actual class	Ture Low	**135**	4	5	144
	Ture Moderate	2	**42**	2	46
	Ture High	7	1	**27**	35
	Total	144	47	34	225

Table 4.70 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting English performance using NB.

Table 4.70: Measures in Predicting English Performance Using NB

Criterion	**Value (%)**
AUC	97.8
Accuracy	90.7
F1-Score	90.6
Precision	90.6
Recall	90.7
MCC	82.2

b) DT

The confusion matrix in table 4.71 shows the results of DT algorithm in predicting English performance. The model correctly classified 199 out of 225 students, resulting in an accuracy of 88.4%. Among low performers, 131 were predicted correctly, with 13 misclassified. For moderate students, 42 were accurately identified, while 4 were predicted as low. In the high-performing group, 26 were correctly classified, with 9 misclassified as low or moderate. Misclassifications mainly occurred between adjacent levels.

Table 4.71: Confusion Matrix in Predicting English Performance Using DT

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
Actual class	Ture Low	**131**	12	1	144
	Ture Moderate	4	**42**	0	46
	Ture High	8	1	**26**	35
	Total	143	55	27	225

Table 4.72 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting English performance using DT.

Table 4.72: Measures in Predicting English Performance Using DT

Criterion	Value (%)
AUC	96.4
Accuracy	88.4
F1-Score	88.5
Precision	89.2
Recall	88.4
MCC	78.2

c) ANNs

The confusion matrix in table 4.73 shows the results of ANNs algorithm in predicting English performance. The model correctly predicted 201 out of 225 English performance cases, with an accuracy of 89.3%. Among low-performing students, 136 were correctly classified, while 8 were misclassified. For moderate students, 43 were accurate, and 3 were predicted as low. In the high-performing group, 22 were correctly identified, with 13 misclassified as low or moderate. Most errors occurred in high-level predictions.

Table 4.73: Confusion Matrix in Predicting English Performance Using ANNs

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
Actual class	Ture Low	**136**	6	2	144
	Ture Moderate	3	**43**	0	46
	Ture High	8	5	**22**	35
	Total	147	54	24	225

Table 4.74 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting English performance using ANNs.

Table 4.74: Measures in Predicting English Performance Using ANNs

Criterion	Value (%)
AUC	97.6
Accuracy	89.3
F1-Score	89.0
Precision	89.7
Recall	89.3
MCC	79.6

d) SVMs

The confusion matrix in table 4.75 shows the results of SVMs algorithm in predicting English performance. The model correctly predicted 203 out of 225 students' English performance, with an accuracy of 90.2%. Among low-performing students, 137 were accurately identified, while 7 were misclassified. For moderate performers, 41 were correct, and 5 were misclassified. In the high-performing group, 25 were correctly predicted, with 10 misclassified as low or moderate. Most errors occurred between high and lower levels.

Table 4.75: Confusion Matrix in Predicting English Performance Using SVMs

		Predicted class			
		Predicted Low	Predicted Moderate	Predicted High	Total
Actual class	Ture Low	**137**	5	2	144
	Ture Moderate	4	**41**	1	46
	Ture High	7	3	**25**	35
	Total	148	49	28	225

Table 4.76 lists the AUC, accuracy, F1-Score, precision, recall, and MCC in predicting English performance using SVMs.

Table 4.76: Measures in Predicting English Performance Using SVMs

Criterion	Value (%)
AUC	95.7
Accuracy	90.2
F1-Score	90.0
Precision	90.2
Recall	90.2
MCC	81.1

Table 4.77 summarizes the measures of four DM algorithms in predicting students' English performance.

Table 4.77: The Measures of Four DM Algorithms in Predicting Students' English Performance

DM algorithms	Prediction Measures (%)					
	ACC	Accuracy	F1-Score	Precision	Recall	MCC
NB	97.8	90.7	90.6	90.6	90.7	82.2
DT	96.4	88.4	88.5	89.2	88.4	78.2

Table 4.77 continued

ANNs	97.6	89.3	89.0	89.7	89.3	79.6
SVMs	95.7	90.2	90.0	90.2	90.2	81.1

Table 4.77 shows that NB achieved the highest accuracy, at 90.7%, followed by SVMs (90.2%). ANNs followed with 89.3%, while DT had the lowest Accuracy at 88.4%. This indicates that NB performed best in predicting English performance.

4.7 Relationship between Student NCEE Performance in Chinese, Mathematics and English

RQ3: What is the relationship between student performance in Chinese, Mathematics and English?

In Section 4.5.1, English is the sixth predictor in NB, fourth predictor in DT, and sixth and twelfth predictor in SVM to Chinese. Mathematics is the twelfth predictor in ANNs to Chinese. So, from Section 4.5.1, the rank factors tell us that there are relationships between Chinese and "English, Mathematics". While in section 4.5.2, among the best predictors of the four algorithms, neither Chinese nor English turned out to be the best factor for predicting Mathematics performance. Therefore, for the time being, no relationship between Mathematics and "Chinese, English" can be observed from Section 4.5.2. In section 4.5.3, Chinese is the thirteenth predictor in DT, and twelfth and seventeenth predictor in SVMs to English. So, from Section 4.5.3, the rank factors tell us that there is relationship between English and Chinese.

To explore the exact relationship among Chinese, Mathematics, and English performance, this study used IBM SPSS statistics software (version 30.0.0.0(172)) for analysis. The specific data analysis method is Pearson's correlation coefficient. The results are shown in Table 4.78.

Table 4.78: The Results of Relationship among Chinese, Mathematics, and English Performance

Assessment	Pearson r values	Assessment	Pearson r values	Assessment	Pearson r values
1 Chinese and 1 Mathematics	.124**	1 Chinese and 1 English	.308**	1 Mathematics and 1 English	.220**
2 Chinese and 2 Mathematics	.264**	2 Chinese and 2 English	.377**	2 Mathematics and 2 English	.402**
3 Chinese and 3 Mathematics	.220**	3 Chinese and 3 English	.311**	3 Mathematics and 3 English	.368**
4 Chinese and 4 Mathematics	.267**	4 Chinese and 4 English	.416**	4 Mathematics and 4 English	.365**
5 Chinese and 5 Mathematics	.197**	5 Chinese and 5 English	.404**	5 Mathematics and 5 English	.377**

**Correlation is significant at the 0.01 level (2-tailed)

In Table 4.78, ** marks indicate that the correlation coefficient is significant at the $p < 0.01$ level. A strong correlation ($r > 0.50$) means a strong positive linear relationship exists between two variables. A moderate correlation ($0.30 < r \leqslant 0.50$) shows a moderate positive relationship. A weak correlation ($r \leqslant 0.30$) indicates a weak positive linear relationship. See Appendix E for details.

Chinese and Mathematics correlations remained uniformly weak across all five assessments, with Pearson r values of .124 in assessment 1, .264 in assessment 2, .220 in assessment 3, .267 in assessment 4, and .197 in assessment 5. None of these coefficients reached the threshold for a moderate association ($r > .30$), indicating that Chinese performance contributes minimally to the prediction of Mathematics performance at any measurement point. This result is consistent with ANNs result in which Mathematics is only the twelfth predictor to Chinese, which indicates a weak relationship.

Chinese and English correlations were consistently moderate in every assessment. The observed r values were .308 in assessment 1, .377 in assessment 2, .311 in assessment 3, .416 in assessment 4, and .404 in assessment 5. Each value falls within the $.30 < r \leqslant .50$ range, reflecting a stable and meaningful overlap in the linguistic skills underpinning both Chinese and English performance throughout the study period. This result is consistent with the data mining results in which English is the sixth predictor in NB, fourth predictor in DT, and sixth and twelfth predictor in SVM to Chinese.

Mathematics and English correlations began as weak but strengthened to moderate levels after the first assessment. Specifically, r = .220 in assessment 1, followed by r = .402 in assessment 2, r = .368 in assessment 3, r = .365 in assessment 4, and r = .377 in assessment 5. Thus, whereas the initial link between quantitative reasoning and language proficiency was limited, subsequent measurements all met the criterion for moderate association.

4.8 Recommendation of Subject Options for Each Student Using DT Rules

This section examines the characteristics of students achieving high performance in six optional subjects, Physics, Chemistry, Biology, History, Politics, and Geography using DT rules.

Under the current NCEE framework, senior secondary students must commit to one of twelve prescribed subject options: Option 1 – Physics, Chemistry, and Biology; Option 2 – Physics, Chemistry, and Politics; Option 3 – Physics, Chemistry, and Geography; Option 4 – Physics, Biology, and Politics; Option 5 – Physics, Biology, and Geography; Option 6 – Physics, Politics, and Geography; Option 7 – History, Politics, and Geography; Option 8 – History, Politics, and Chemistry; Option 9 – History, Politics, and Biology; Option 10 – History, Geography, and Chemistry; Option 11 – History, Geography, and Biology; and Option 12 – History, Biology, and Chemistry.

Selecting the most suitable trio is critical because it shapes teaching allocation, examination content, and university admission prospects. This subsection therefore converts the DT rules extracted for Physics, Chemistry, Biology, History, Politics, and Geography into practical counselling guidance. Each rule pinpoints the blend of academic indicators, psychological attributes, and contextual supports that most reliably predict high achievement in a given subject. By aligning every student's profile with these predictors, the analysis recommends the single option whose three-subject constellation maximises the probability of overall excellence while respecting individual strengths and constraints. The resulting advice is evidence-based, transparent, and immediately actionable for schools, students, and families.

By identifying these factors, this study aims to provide targeted recommendations of subject options for each student based on their characteristics, helping them achieve optimal academic performance.

4.8.1 Predicting Student NCEE Performance in Physics

RQ4a: What are the characteristics of students with high performance in Physics using DT rules?

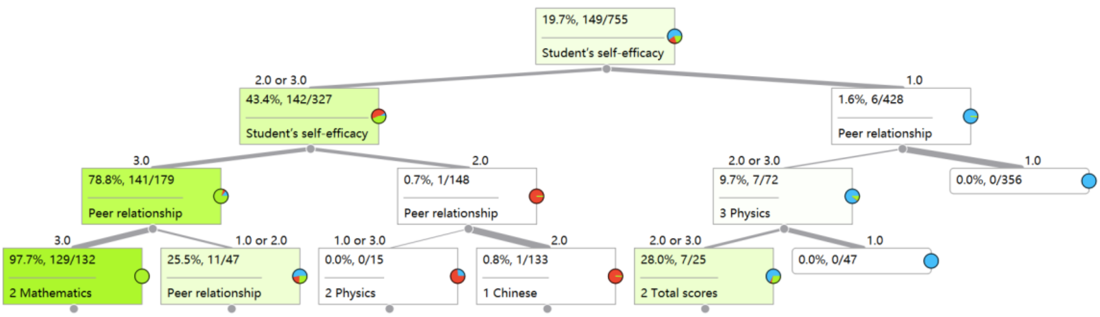

Figure 4.13: High Performance in Physics Using DT Rules

Figure 4.14 shows the characteristics of students achieving high performance in Physics using DT rules. This involves 149 out of 755 students (19.7%). These students are primarily influenced by student's self-efficacy and peer relationship. A significant determinant of high Physics performance in NCEE is student's self-efficacy. Among moderate or high self-efficacy level students, 43.4% (142 out of 327) went on to achieve high Physics performance. Within this subset, high self-efficacy level students, 78.8% (141 out of 179) achieved success. These high self-efficacy level students were further influenced by peer relationship. Students with high level of peer relationship (97.7%, 129 out of 132) showed a strong likelihood of maintaining high performance.

The DT highlights that high student's self-efficacy and high level of peer relationship significantly contribute to success. So the rule is as follows:

IF (Student's Self-efficacy = High) AND (Peer Relationship = High) THEN 2025 NCEE Performance in Physics = High.

4.8.2 Predicting Student NCEE Performance in Chemistry

RQ4b: What are the characteristics of students with high performance in Chemistry using DT rules?

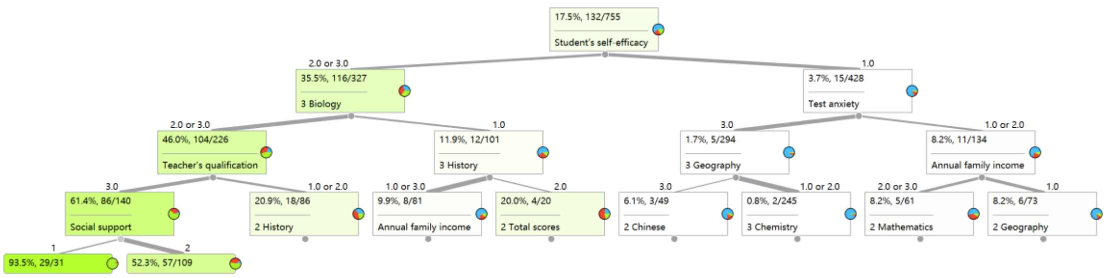

Figure 4.14: High Performance in Chemistry Using DT Rules

Figure 4.15 shows the characteristics of students achieving high performance in Chemistry using DT rules. This involves 132 out of 755 students (17.5%). These students are also primarily influenced by student's self-efficacy, past Biology performance, teacher's qualification, and social support. A significant determinant of high Chemistry performance in NCEE is also students' self-efficacy. Among these students who have moderate or high level of self-efficacy, 35.5% (116 out of 327) went on to achieve high Chemistry performance. One interesting finding is that students (46.0%, 104 out of 226) with moderate or high Biology performance in final-term examination before subject options in 2022 also displayed notable success in Chemistry. These students with moderate or high Biology in final-term examination before subject options in 2022 were further influenced by teacher's qualification. Students with high teacher's qualification (61.4%, 86 out of 140) still maintain high performance. Within this subset, social support played a critical role, with 93.5% (29 out of 31) of students with social support achieving more success.

The DT highlights that high student's self-efficacy, foundational knowledge in Biology, high level of teacher's qualification, and more social support significantly contribute to success. So the rule is as follows:

IF (Student's Self-efficacy = Moderate or High) AND (Biology Performance in Final-term Examination Before Subject Options in 2022 = Moderate or High) AND (Teacher's Qualification = High) AND (Social Support = Yes) THEN 2025 NCEE Performance in Chemistry = High.

4.8.3 Predicting Student NCEE Performance in Biology

RQ4c: What are the characteristics of students with high performance in Biology using DT rules?

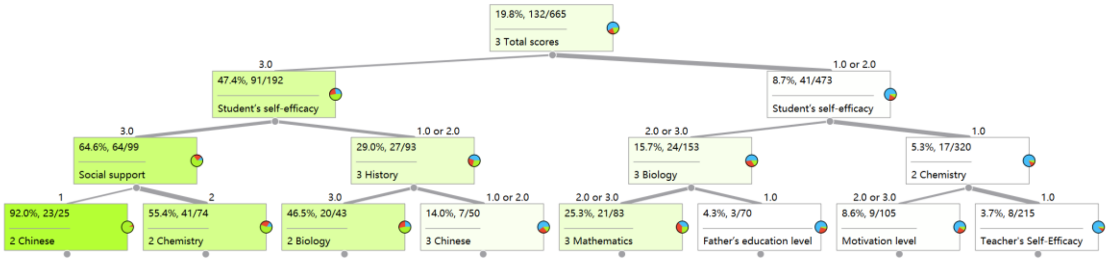

Figure 4.15: High Performance in Biology Using DT Rules

Figure 4.16 shows the characteristics of students achieving high performance in Biology using DT rules. This involves 132 out of 665 students (19.8%). These students are primarily influenced by total scores in final-term examination before subject options in 2022, student's self-efficacy, and social support. A significant determinant of high Biology performance in NCEE is total scores in final-term examination before subject options in 2022. Among these students who have high total scores, 47.4% (91 out of 192) went on to achieve high Biology performance. These students with high total scores in final-term examination before subject options in 2022 are further influenced by student's self-efficacy. Students with high self-efficacy (64.6%, 64 out of 99) still maintain high performance. Within this subset, students with social support (92.0%, 23 out of 25) showed a likelihood of maintaining high performance.

The DT highlights that high total scores, high student's self-efficacy, and social support significantly contribute to success. So the rule is as follows:

IF (Total Scores in Final-term Examination before Subject Options in 2022 = High) AND (Student's Self-efficacy= High) AND (Social Support = Yes) THEN 2025 NCEE Performance in Biology = High.

4.8.4 Predicting Student NCEE Performance in History

RQ4d: What are the characteristics of students with high performance in History using DT rules?

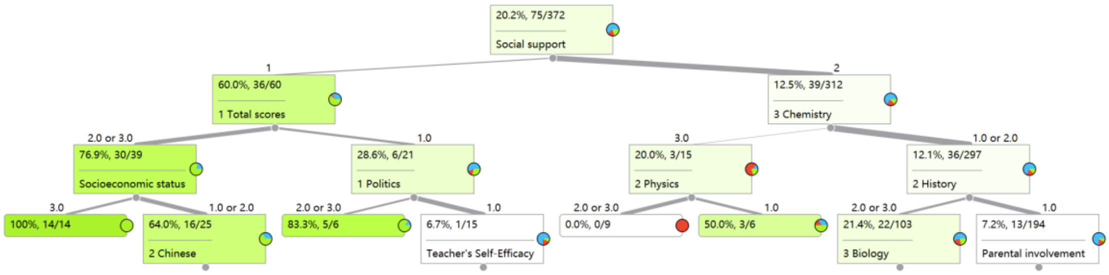

Figure 4.16: High Performance in History Using DT Rules

Figure 4.17 shows the characteristics of students achieving high performance in History using DT rules. This involves 75 out of 372 students (20.2%). These students are primarily influenced by past performance and peer relationships. A significant determinant of high History performance in NCEE is social support. Among these students who received social support, 60.0% (36 out of 60) went on to achieve high performance in NCEE. This group with moderate or high total scores in enrolment performance in 2022 (76.9%, 30 out of 39) displayed notable success. Among this group who had high socioeconomic status (100%, 14/14) all achieved high performance. Even if the total scores in enrolment performance in 2022 were poor (28.6%, 6/21), as long as the Politics enrolment performance in 2022 were moderate or high, these students (83.3, 5/6) could still achieve high performance in History in NCEE.

The DT highlights that social support, past total scores, socioeconomic status, and past Politics performance significantly contribute to success. So the rules are as follows:

IF (Social Support = Yes) AND (Total Scores in Enrolment Performance in 2022 = Moderate or High) AND (Socioeconomic Status = High) THEN NCEE Performance in History = High.

IF (Social Support = Yes) AND (Total Scores in Enrolment Performance in 2022 = Low) AND (Politics Enrolment Performance in 2022 = Moderate or High) THEN 2025 NCEE Performance in History = High.

4.8.5 Predicting Student NCEE Performance in Politics

RQ4e: What are the characteristics of students with high performance in Politics using DT rules?

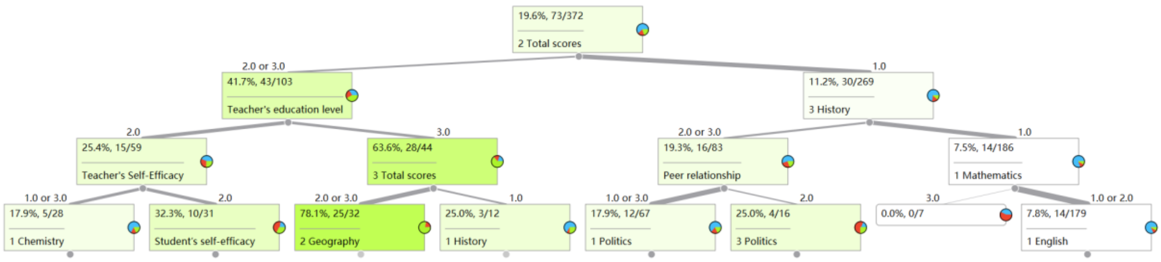

Figure 4.17: High Performance in Politics Using DT Rules

Figure 4.18 shows the characteristics of students achieving high performance in Politics using DT rules. This involves 73 out of 372 students (19.6%). These students are primarily influenced by past total scores, and teacher's education level. A significant determinant of high performance in Politics in NCEE is past total scores. Among these high-performing students, 41.7% (43 out of 103) went on to achieve high performance in NCEE. These students are further influenced by teacher's education level. Students with a high level of teacher's education level (63.6%, 28 out of 44) achieved success. Among students with a high level of teacher's education level, total score in final-term examination before subject options in 2023 as a distinguishing factor. Students with high past total scores achieved a fairish success rate of 78.1% (25 out of 32).

The DT highlights that past total scores, and teacher's education level. So the rule is as follows:

IF (Total Score in Mid-term Examination before Subject Options in 2022 = Moderate or High) AND (Teacher's Education Level = High) AND (Total Score in Final-term Examination before Subject Options in 2023 = Moderate or High) THEN 2025 NCEE Performance in Politics = High.

4.8.6 Predicting Student NCEE Performance in Geography

RQ4f: What are the characteristics of students with high performance in Geography using DT rules?

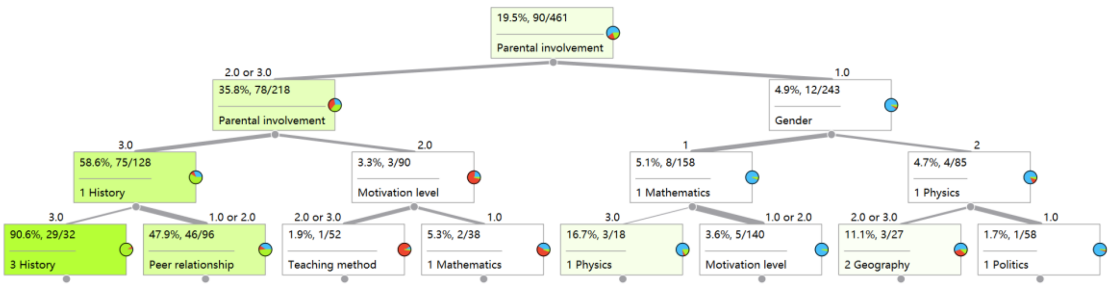

Figure 4.18: High Performance in Geography Using DT Rules

Figure 4.19 shows the characteristics of students achieving high performance in Geography using DT rules. This involves 90 out of 461 students (19.5%). These students are primarily influenced by parental involvement in student's learning and past History performance. A significant determinant of high performance in Geography performance is parental involvement in student's learning. Among moderate or high-performing Geography students, 35.8% (78 out of 218) went on to achieve high performance in NCEE. Students with high-level of parental involvement in student's learning (58.6%, 75 out of 128) showed a strong likelihood of maintaining high performance. Within this subset, History enrolment performance in 2022 played a critical role, with 90.6% (29 out of 32) of students with high Geography performance in NCEE achieving success.

The DT highlights that parental involvement in student's learning and foundational knowledge in History significantly contribute to success. So the rule is as follows:

IF (Parental involvement in student's learning = High) AND (History enrolment performance in 2022 = High) THEN 2025 NCEE Performance in Geography = High.

4.8.7 The Characteristics of High-performing Students for Each Subject

This section explores the DT-based rules that identify key characteristics of high-performing students across six NCEE subjects. By examining academic, psychological, and contextual factors, this study uncovers the characteristics of high-performing students for each subject (see Table 4.79).

Table 4.79: The Characteristics of High-performing Students for Each Subject

Subject	Characteristics of high-performing students
Physics	High student's self-efficacy, high peer relationship
Chemistry	Moderate/High student's self-efficacy, Moderate/High Biology final-term performance (2022), High teacher's qualification, social support
Biology	High final-term total scores (2022), High student's self-efficacy, Social support
History	Social support, Moderate/High enrolment total scores (2022), High socioeconomic status; or Low enrolment total scores (2022) + Moderate/High enrolment Politics performance
Politics	Moderate/High mid-term total scores (2022), High teacher's education level, Moderate/High final-term total scores (2023)
Geography	High parental involvement in student's learning, High enrolment History performance (2022)

4.9 Chapter Summary

In this chapter, individual, family, school, and social factors that affect and predict performance were obtained from the school database and teacher-student surveys. Most research questions were analysed using four algorithms in Orange software: NB, DT, ANNs, and SVMs. The most important factors to predict NCEE performance in Chinese, Mathematics, and English have been identified, and the highest accuracy predictive model has also been evaluated. The relationships between Chinese, Mathematics, and English performance were analysed. The rules for high performance in Physics, Chemistry, Biology, Politics, History, and Geography have also been analysed. The next chapter is the discussion, which explains the results.

CHAPTER 5

DISCUSSION

5.1 Introduction

This chapter discusses the results of Chapter 4 and, in connection with the 11 research questions of this study, successfully achieves all the research objectives of this study.

5.2 Discussion of the Results

5.2.1 The Most Important Factors to Predict NCEE Performance in Chinese, Mathematics, and English

5.2.1.1 About Chinese Subject

This study analysed the most important factors to predict student NCEE performance in Chinese using the NB algorithm. The findings showed that motivation level (weight 0.845) is the most critical factor, followed by teaching method (weight 0.110), gender (weight 0.094), 4 Chinese (weight 0.087), teacher's self-efficacy (weight 0.076), 4 English (weight 0.068), 3 Chinese (weight 0.065), 2 Politics (weight 0.056), 2 Biology (weight 0.056), and teacher's qualification (weight 0.052).

Motivation level was the strongest predictor. Motivation level robustly predicts students' Chinese performance by capturing learners' intrinsic drive and self-regulatory capacities essential for language mastery. Motivated students allocate greater cognitive resources to vocabulary expansion, character writing, and text comprehension, yielding sustained improvements across reading, writing, and oral assessments through enhanced learning efficiency, clear goal setting, and perseverance (Jiang, 2023). Intrinsic motivation further bolsters self-efficacy for self-regulation, which mediates writing performance by promoting strategic planning, persistence in revision, and adaptive study strategies (Li et al., 2024). Empirical evidence indicates a strong positive correlation between reading motivation and reading comprehension scores ($r = .720$, $p < .001$), with motivation accounting for over 50% of variance in comprehension outcomes (Ma & Zhao, 2025). Motivation level is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Teaching method serves as a predictor of students' Chinese performance by directly shaping instructional engagement and skill acquisition. Flipped classroom models, which shift foundational content delivery outside class and dedicate in-person time to collaborative problem solving, yielded significant gains in classical Chinese reading comprehension, with treatment groups outperforming controls by large effect sizes ($d > .80$) after a two-year intervention (Lau & Qian, 2024). Project-based learning (PjBL), involving authentic, sustained writing tasks, produced a 0.68 correlation with improved argumentative composition quality, mediated by enhanced metacognitive strategy use and raised overall writing scores significantly over six weeks of instruction (Gao et al., 2024). Cooperative learning approaches, characterized by structured peer interaction through conversation cards, role-play, storytelling, and debate, have been shown to foster foreign-language expression skills and boost communicative fluency in Chinese as a foreign language classroom (Qorie et al., 2024). These evidence-based methods, each prioritizing active learner engagement, authentic tasks, and metacognitive reflection, account for substantial variance in reading, writing, and speaking outcomes. Integrating such teaching methods within formative assessment frameworks enables educators to predict performance trajectories accurately and tailor pedagogical interventions to sustain long-term gains in Chinese performance. Teaching method is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Gender exerts a consistent influence on Chinese performance, with girls outperforming boys across key literacy domains. In large-scale assessments of lower-secondary students, girls scored approximately 0.5 SD higher in Chinese after controlling for hukou status, cognitive ability, and school effects, underscoring gender as a robust predictor of achievement (Luo et al., 2021). Early language trajectories also favour females: peri-urban girls aged 18 - 24 months demonstrate significantly higher receptive and expressive language scores, reflecting enriched conversational environments that lay the groundwork for later literacy advantages (Ma et al., 2024). Across reading mediums, Chinese undergraduates display positive performance trends for females, who consistently achieve higher mean scores than males in both print and smartphone-based comprehension tests, suggesting gendered patterns of engagement and strategy use in digital and traditional contexts (Tian et al., 2024). Mechanistically, girls engage more frequently in reading activities, apply metacognitive strategies such as planning and self-monitoring, and report stronger linguistic self-efficacy, collectively enhancing vocabulary acquisition, character retention, and composition quality. Incorporating gender into predictive models thus refines early-warning systems for Chinese performance and informs targeted interventions, such as peer model reading groups for boys and metacognitive strategy workshops, that aim to narrow gender gaps and foster equitable literacy outcomes. Gender is also the best predictor for the two algorithms: DT, and ANNs.

Past Chinese performance serves as a powerful predictor of subsequent Chinese performance because it captures the cumulative mastery of foundational skills, character recognition, vocabulary breadth, and syntactic fluency that scaffold later literacy development. Early Chinese scores exhibit high autoregressive stability, for instance, first grade character reading accuracy explains over 60% of variance in third grade reading comprehension (Lin et al., 2019). At age 11, syntactic awareness uniquely forecasts writing composition quality one year later, highlighting how prior performance in linguistic tasks precedes expressive competence (Tong & McBride, 2016). Moreover, historical Chinese grades exert cascading effects on academic self-perceptions, which then fuel motivation and strategic learning behaviours, collectively reinforcing achievement trajectories across elementary grades (Fu et al., 2020). These findings indicate that past performance not only reflects existing skill levels but also shapes affective and metacognitive factors that drive future success. Incorporating longitudinal Chinese scores into predictive models refines early-identification systems for underachievement risk and supports the design of targeted interventions, such as precision tutoring in morphological compounding and syntax drills that bolster long-term performance. Furthermore, tailoring feedback based on error patterns in past assessments enables adaptive remediation addressing persistent gaps in character usage and composition structure. This evidence underscores the centrality of longitudinal performance metrics in both research and practical application within Chinese education. Past 4 Chinese performance is also the best predictor for the two algorithms: DT, and ANNs. While past 3 Chinese performance is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Teacher's self-efficacy strongly predicts students' Chinese performance by shaping instructional quality, classroom management, and teacher-student interactions. In Chongqing secondary schools, higher teacher self-efficacy, especially regarding classroom management, correlated significantly with middle-school students' Chinese scores, accounting for over 40% of performance variance (Zhang et al., 2024). High-efficacy teachers design more coherent lessons and provide clearer explanations, which enhance students' character recognition and reading comprehension (Song et al., 2025). Moreover, teachers confident in their pedagogical abilities employ diverse formative assessments and timely feedback, fostering students' metacognitive monitoring and self-regulated learning, key drivers of writing and oral proficiency (Jerrim et al., 2025). Empirical evidence indicates that self-efficacious teachers sustain positive classroom climates, reducing anxiety and increasing student engagement in Chinese tasks, thereby elevating test performance across reading, writing, and listening domains (Jerrim et al., 2025). Integrating teacher self-efficacy measures into educational evaluations enables early identification of instructional gaps and informs targeted professional development that further amplifies efficacy beliefs and, consequently, student performance. These findings underscore the pivotal role of teacher self-efficacy in both predicting and enhancing Chinese learning performance across varied educational settings. Teacher's self-efficacy is also the best predictor for the three algorithms: DT, ANNs, and

SVMs.

Past English performance predicts students' Chinese performance by indexing transferable metalinguistic skills that underpin literacy across languages. In Chinese-English bilingual fourth graders, past English word reading fluency and linguistic comprehension together explained over 50% of variance in Chinese reading comprehension, with word reading holding greater predictive weight in bilingual contexts (Zang et al., 2024). Neuroimaging evidence further demonstrates that higher English proficiency predicts neural patterns of phonological processing in Chinese, reflecting L2-to-L1 transfer: bilingual children with stronger English skills exhibited bilateral frontal activation during Chinese phoneme tasks, which correlated with superior Chinese reading scores (Kou et al., 2024). Moreover, a cross-cultural meta-analysis confirmed moderate and significant correlations ($r \approx .49$) between morphological awareness, often developed through English instruction and Chinese reading comprehension in L2 learners, highlighting morphology as a key cross linguistic resource (Ke, 2025). These findings indicate that past English performance captures foundational decoding strategies, phonological agility, and morphological insight that transfer to Chinese, thereby making historical English performance a powerful predictor in early warning and tailored intervention models for Chinese development. Past English performance is also the best predictor for the two algorithms: DT, and SVMs.

Past performance in ideological and political education (hereafter Politics) constitutes a predictor of students' Chinese performance by honing critical reading, argumentation, and discourse skills transferable to language tasks. High achievers in Politics demonstrate superior text analysis abilities and rhetorical structuring, which directly enhance comprehension and composition in Chinese assessments (Yang et al., 2024). Moreover, students excelling in Politics courses often engage in extensive essay writing on socio political topics, fostering metacognitive monitoring and precision in character usage that correlate strongly with elevated Chinese writing scores (Wang et al., 2023). Empirical evidence also shows that the linguistic framing and argumentative strategies cultivated through Politics curricula improve students' narrative cohesion and argumentative coherence in Chinese essays, accounting for significant variance in overall Chinese grades (Sang, 2024). These cross-disciplinary transfers underscore how Politics performance not only reflects domain-specific knowledge but also shapes cognitive and compositional frameworks essential for Chinese literacy. Incorporating historical Politics performance into predictive models thus enhances early identification of students at risk of underachievement in Chinese, enabling targeted interventions, such as structured debate workshops and integrated writing clinics that leverage Politics derived competencies to elevate Chinese performance. Past Politics performance is not the best predictor for the other three algorithms.

Past Biology performance predicts Chinese performance by indexing transferable literacy and cognitive skills honed in scientific study. Science learning activities, such as interpreting technical texts, analyzing diagrams, and constructing written explanations, engage deep processing and self-regulated strategies that directly bolster reading comprehension, question generation, and knowledge building (Cano et al., 2014). Moreover, the central executive demands of Biology tasks, planning experiments and integrating multimodal information, parallel the working-memory updating processes that uniquely predict Chinese reading comprehension among primary students (Gao et al., 2023). Vocabulary and morphological awareness cultivated through rigorous Biology terminology acquisition also correlate with gains in Chinese character reading, explaining significant variance in literacy outcomes (Zang et al., 2024). Finally, metacognitive reading strategies practiced when engaging dense scientific passages, such as self-monitoring, summarizing, and inferencing, mediate the relationship between science performance and Chinese composition quality, underscoring the role of strategic reading in writing coherence (Liao et al., 2020). Collectively, these findings demonstrate that past Biology performance encapsulate the cognitive, metacognitive, and linguistic foundations essential for successful Chinese performance. Past Biology performance is also not the best predictor for the other three algorithms.

Teacher's qualification, encompassing formal certification and depth of subject-matter knowledge, emerges as a predictor of students' Chinese performance by underpinning instructional coherence and fostering strategic literacy practices. Teachers who complete specialized reading-teacher competence certification programs demonstrate stronger pedagogical content knowledge and instructional design skills, which translate into significant gains in students' character recognition and reading comprehension (Yuan & Xu, 2024). Moreover, teachers' language and literacy knowledge, a core facet of professional qualification, reliably forecasts students' improvements in foundational reading skills, accounting for substantial variance in early Chinese reading outcomes (Porter et al., 2024). In rural Chinese contexts, higher-qualified teachers exhibit greater trust and commitment to instructional goals, mediating the influence of school leadership on students' reading literacy and explaining over 40% of performance differences (Zhu et al., 2022). These qualified educators employ diverse formative assessments and metacognitive strategy support, fostering students' self-regulated learning that directly enhances Chinese writing quality and comprehension scores. Integrating teacher qualification metrics into predictive models thus refines the early identification of students at risk and informs targeted professional development interventions, such as advanced certification programs and content-focused workshops that can elevate long-term Chinese performance. Teacher's qualification is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

For RQ1a, this study identifies the most important predictors of Chinese NCEE performance using the NB algorithm. Motivation level emerged as the strongest factor, significantly influencing reading, writing, and comprehension outcomes. Teaching method also played a critical role, with evidence supporting flipped classrooms and project-based learning. Gender differences favored girls, linked to stronger metacognitive strategies and language engagement. Past Chinese and English performance were strong indicators, reflecting skill accumulation and cross-linguistic transfer. Politics and Biology performance predicted performance via transferable cognitive and literacy skills. Teacher-related factors, including self-efficacy and qualification, shaped instructional quality and student's Chinese performance.

These findings align with Bronfenbrenner's Ecological Systems Theory by illustrating how multiple layers influence academic performance. Motivation and teaching methods operate at the microsystem (classroom level), while gender differences and past performance reflect mesosystem interactions (e.g., school-family links) and macrosystem cultural norms. However, limitations include the cross-sectional design, which limits causal inferences, and potential sample bias in Chinese contexts. For policy, educational departments should promote teacher training on motivational strategies and gender-sensitive pedagogy, while practitioners can foster home-school partnerships to enhance literacy engagement, addressing ecological factors holistically.

5.2.1.2 About Mathematics Subject

The current study identified test anxiety (weight 0.827), father's education level (weight 0.579), mother's education level (weight 0.543), socioeconomic status (weight 0.419), peer relationship (weight 0.391), student's self-efficacy (weight 0.344), 2 Mathematics (weight 0.192), teaching method (weight 0.134), 4 Mathematics (weight 0.132), 3 Physics (weight 0.123), teacher's self-efficacy (weight 0.118), 3 Chemistry (weight 0.112), and 4 total scores (weight 0.100) as significant factors predicting student NCEE performance in Mathematics using NB algorithm.

Test anxiety robustly predicts middle school students' Mathematics performance by undermining working memory resources and increasing maladaptive avoidance behaviours during assessments. A large-scale meta analysis of nearly one million participants confirmed that test anxiety is negatively correlated with math achievement ($r = - 0.23$), indicating that higher anxiety reliably forecasts lower Mathematics scores (Caviola et al., 2021). Secondary school data show that approximately one in six students experiences high test anxiety, and those students score significantly below their low anxiety peers on standardized Mathematics tests (Putwain & Daly, 2014). Although mathematical anxiety exerts a stronger overall effect, test anxiety independently explains unique variance in performance, predicting poorer outcomes even after controlling for prior achievement and general anxiety (Namkung et al., 2019). In middle school cohorts, test anxiety correlates positively with math anxiety and inversely with grades, with moderate to large effect sizes across studies (Arslan, 2020). These findings suggest that including test anxiety measures in predictive models enhances identification of students at risk and guides interventions, such as cognitive behavioural coping skills training and test taking strategy workshops that target anxiety reduction to improve Mathematics performance. Test anxiety is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Parental education level robustly predicts middle school students’ Mathematics performance by reflecting parents’ capacity to provide cognitive stimulation, educational resources, and supportive home learning environments. Higher parental educational attainment correlates with stronger parental beliefs in the value of Mathematics, which fosters enhanced motivation and achievement in children (Hidayatullah & Csíkos, 2024). Multilevel analysis reveals that each increase in parents’ degree level is associated with significant gains in students’ math scores, accounting for nearly one fifth of between student variance (Li et al., 2021). Parental education also shapes homework support quality, where educated parents more effectively scaffold problem solving strategies and monitor children’s learning processes, leading to improved working memory and conceptual understanding in Mathematics (Pang et al., 2013). Longitudinal data further show that parental education exerts indirect effects on mathematics outcomes through children’s self-efficacy and persistence, which mediate the relationship between background factors and achievement (Davis - Kean, 2005). These findings indicate that integrating parents’ education metrics into predictive models enhances early identification of students at risk and guides interventions, like parent workshops on effective math support that leverage parental expertise to elevate Mathematics performance. Parental education level is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Socioeconomic status (SES) reliably predicts middle school students' Mathematics performance by capturing access to educational resources, home learning support, and school quality that collectively foster cognitive and motivational foundations for math learning. Multilevel analysis of 10,784 Chinese middle school students demonstrated that both family-level and school-level SES are positively associated with individual Mathematics scores, with school context amplifying family advantages in ruraierarchical regression using national panel data confirmed that higher parental education and household income predict significant gains in students' math achievement, accounting for meaningful between-student variance after controlling for prior performance and school factors (Xu, 2023). Cross-national PISA 2022 comparisons found that economic, social, and cultural status are among the strongest predictors of math performance in Singapore, Korea, Finland, and Denmark, explaining over twenty six percent of student-level variance across contexts (Niu et al., 2025). An umbrella review of 48 meta-analyses further synthesized that the SES achievement association in Mathematics represents a moderate effect size and is mediated by factors such as parental involvement, school instructional quality, and student motivation (Tan, 2024). Integrating SES metrics into predictive models thus enhances early identification of students at risk and informs targeted equity interventions to boost Mathematics performance. SES is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Peer relationship predicts middle school students' Mathematics performance by fostering collaborative problem solving, sustained engagement, and positive academic norms. Peer relationships directly enhance academic outcomes and indirectly boost mathematics scores through increased learning motivation and engagement, as evidenced by chain mediating effects in junior high learners (Shao et al., 2024). Quasi experimental research in Chinese middle schools shows that classmates' prior academic performance raises individual Mathematics achievement by up to 0.17 standard deviations via study time and confidence gains (Chen et al., 2023). Reciprocal peer tutoring interventions yield medium to large effect sizes (Hedges' g = 0.48) for Mathematics achievement, demonstrating that well-structured peer support enhances conceptual understanding and self-efficacy (Moliner & Alegre, 2022). Experimental studies also report that same age peer tutoring programs produce over thirteen percent improvements in Mathematics test scores within weeks of implementation, underlining the potency of guided peer interactions (Moliner & Alegre, 2020). These findings indicate that integrating peer relationship metrics into predictive models sharpens early identification of students at risk and informs classroom practices that leverage peer dynamics to elevate Mathematics performance. Peer relationship is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Student's self-efficacy in Mathematics consistently predicts middle school students' Mathematics performance because it captures confidence in problem solving, persistence under challenge and the deployment of effective learning strategies. A scoping review found strong associations between Mathematics self-efficacy and both performance and engagement across diverse student populations (Street et al., 2024). Longitudinal studies reveal reciprocal relations between self-efficacy and achievement, with prior self-efficacy forecasting later gains in Mathematics scores and vice versa (Arens et al., 2022). Empirical research shows that self-efficacy mediates the link between previous attainment and future performance by fostering mastery orientation and sustained effort, accounting for unique variance beyond prior test scores (Du et al., 2021). Studies of parental support further demonstrate that higher self-efficacy leads to greater engagement and enjoyment in Mathematics tasks, which in turn elevate test outcomes among adolescents (Sağkal & Sönmez, 2022). These findings indicate that incorporating self-efficacy measures into predictive models improves the identification of students at risk and informs interventions that strengthen efficacy beliefs and thereby enhance Mathematics performance. Student's self-efficacy is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Past Mathematics performance stands as a predictor of subsequent Mathematics performance by encapsulating cumulative mastery of procedural fluency, conceptual understanding, and effective study strategies. Longitudinal machine learning analyses of German secondary students reveal that prior year standardized Mathematics scores consistently outpace all other cognitive, motivational, and contextual factors in forecasting next year achievement, with their predictive weight increasing over time (Lavelle‐Hill et al., 2024). A longitudinal meta-analysis of over 58,000 students from kindergarten through grade 12 further confirms that early numeracy skills, a component of past Mathematics performance, predict later Mathematics outcomes with an average correlation of .49, demonstrating a snowball effect on advanced problem solving and word problem performance (Liu et al., 2025). Sequential growth curve research indicates that kindergarten number competence, another facet of early Mathematics mastery, significantly predicts both growth rate and achievement level in Mathematics through third grade, underscoring the enduring impact of past performance (Jordan et al., 2009). These findings suggest that integrating historical Mathematics performance into predictive models enables precise early identification of students at risk and guides targeted interventions, such as tailored fluency drills and scaffolded problem solving workshops that leverage existing strengths to enhance future Mathematics learning. Past Mathematics performance is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Teaching method predicts middle school students' Mathematics performance by structuring learning experiences that engage cognitive processes essential for numeracy. Flipped classroom implementation where content delivery occurs outside class and in person time focuses on problem solving yields significant gains in achievement and interest across diverse settings, with learners in flipped environments outperforming peers by large effect sizes (Egara & Mosimege, 2024). Reciprocal peer tutoring embeds collaborative practice and metacognitive dialogue, producing medium effect sizes (Hedges' $g \approx 0.48$) and promoting both skill mastery and positive attitudes toward Mathematics (Moliner & Alegre, 2022). Technology enhanced methods, such as GeoGebra supported 5E learning models, improve conceptual understanding and retention in geometry, translating to higher test scores and sustained engagement (Uwurukundo et al., 2024). These findings demonstrate that teaching methods which combine active learning, social interaction, and digital tools capture critical dimensions of student cognition and motivation. Integrating measures of instructional approach into predictive models thus refines identification of students at risk and guides targeted professional development such as training in flipped pedagogy, structured peer tutoring protocols, and technology integration workshops to elevate Mathematics performance. Teaching method is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Teacher's self-efficacy in Mathematics predicts middle school students' Mathematics performance because confident teachers employ adaptive instructional strategies, create supportive learning environments, and stimulate student engagement, all of which directly influence achievement. Structural equation modelling in Ghanaian high schools shows that teacher efficacy positively moderates the relationship between students' perception of Mathematics and their achievement outcomes (Akendita et al., 2024). Evidence from the TALIS video study using TIMSS 2015 data indicates that higher teacher self-efficacy aligns with elevated student reported lesson quality and consequent gains in Mathematics scores (Jerrim et al., 2025). Mediational analyses reveal that teacher-student relationship quality enhances mathematics performance through teacher efficacy, highlighting its role in leveraging social dynamics for academic benefit (Hajovsky et al., 2020). Correlational research in South African secondary schools finds that teachers with stronger self-efficacy beliefs implement more effective formative assessment, scaffold complex problem solving, and maintain higher classroom engagement, leading to superior student outcomes in Mathematics (Olawale & Hendricks, 2024). These findings demonstrate that incorporating teacher self-efficacy metrics into predictive models refines early identification of at-risk classrooms and guides professional development initiatives focused on mastery experiences, reflective practice, and collaborative learning communities to elevate Mathematics performance. Teacher's self-efficacy is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Past Physics and Chemistry performance predict middle school students' Mathematics performance because both science disciplines cultivate analytical reasoning, quantitative problem solving, and the strategic application of mathematical concepts that transfer directly to Mathematics tasks. In West African secondary schools, Physics scores correlate with further Mathematics achievement ($r = .68$), indicating that mastery of physics concepts forecasts gains in advanced mathematics topics (Badmus & Jita, 2023). Chemistry performance likewise forecasts mathematics outcomes: a Guilin secondary school study found a robust linear relationship ($r = .74$) between chemistry and mathematics scores, reflecting shared demands for algebraic manipulation and numerical interpretation (Anchen & Ying, 2022). Cognitive reflection, measured by the cognitive reflection test, predicts both physics and Mathematics achievement ($\beta = .45$), underscoring shared metacognitive processes across domains (Doz & Sliško, 2024). Systematic review evidence on STEM transitions reveals that early success in science courses supports sustained Mathematics achievement by reinforcing problem-solving strategies and self-regulation (Kaur et al., 2022). These findings demonstrate that integrating historical Physics and Chemistry grades into predictive models enhances early identification of students at risk in Mathematics and informs interdisciplinary interventions, like integrated STEM problem based learning workshops that leverage science to math cognitive transfer to elevate Mathematics performance. Past Physics performance is also the best predictor for the two algorithms: DT, and ANNs. While Past Chemistry performance is also the best predictor for DT algorithm.

Past total scores predict middle school students' Mathematics performance by integrating cumulative mastery of computation, problem solving, and study habits. Longitudinal machine-learning analysis shows that previous year aggregate Mathematics scores explain more variance in next-year achievement than any cognitive or motivational predictor, with predictive weight increasing across grades (Lavelle-Hill et al., 2024). Meta analytic findings indicate that early total achievement measured at school entry forecasts Mathematics trajectories through grade eight, with average correlations of .49 between initial composite scores and later standardized test results (Liu et al., 2025). Growth-curve studies reveal that kindergarten total readiness scores predict growth rates and achievement levels through middle school by encapsulating foundational numeracy and executive-function capacities (Jordan et al., 2009). Hierarchical modelling of STEM participants in US middle schools confirms that past total academic performance accounts for substantial between-student differences in subsequent Mathematics performance after controlling for socioeconomic factors (Cameron et al., 2015). These findings demonstrate that incorporating historical total test scores into predictive models refines early identification of students at risk and informs adaptive interventions targeting fluency practice and strategy mastery to accelerate Mathematics learning. Past 4 total scores is also the best predictor for SVMs algorithm.

For RQ1b, this study identified the most important predictors of NCEE Mathematics performance using the NB algorithm. Test anxiety was the strongest factor, significantly lowering scores by impairing working memory and promoting avoidance. Father's and mother's education levels influenced performance by shaping home learning environments and cognitive support. Socioeconomic status impacted access to resources and school quality. Peer relationship enhanced learning through collaboration and engagement. Student's self-efficacy improved persistence and strategic learning. Past Mathematics performance reflected cumulative skill mastery. Teaching method influenced achievement through active learning and technology use. Past Physics and Chemistry performance predicted performance via transferable reasoning skills. Teacher's self-efficacy improved instruction and classroom engagement. Past total scores captured academic readiness and supported accurate risk identification.

Using the NB algorithm, this study identified test anxiety as the strongest negative predictor of NCEE Mathematics performance, alongside key influencing factors including parental education, socioeconomic status, peer relationships, self-efficacy, past academic performance, teaching methods, and teacher self-efficacy. The results resonate with Bronfenbrenner's theory, as test anxiety and self-efficacy pertain to the microsystem, whereas parental education and socioeconomic status represent the exosystem and macrosystem, shaping home environments and societal resources. Peer relationships highlight the mesosystem's role in learning. Limitations involve the reliance on self-reported data and cultural specificity, which may affect generalizability. Policy implications include integrating mental health support in schools to mitigate test anxiety and designing community-based programs to involve parents. Practitioners should adopt collaborative learning strategies to leverage peer influences and equitable resource allocation.

5.2.1.3 About English Subject

This study analysed the most important factors to predict student NCEE performance in English using the NB algorithm. The findings showed that annual family income (weight 0.605) is the most critical factor, followed by parental involvement in student's learning (weight 0.411), 4 English (weight 0.264), 3 English (weight 0.160), 2 English (weight 0.159), socioeconomic status (weight 0.132), teacher's qualification (weight 0.116), teacher's self-efficacy (weight 0.115), teaching method (weight 0.110), and 4 total scores (weight 0.102).

Annual family income functions as a significant predictor of secondary students' English performance. Higher income provides access to enriched learning resources, supplementary academic support and quality educational environments that enhance language acquisition (Nyamubi, 2019). Families with greater financial capacity facilitate exposure to books, multimedia content and extracurricular programs that foster vocabulary expansion and comprehension skills essential for English proficiency (Abbasian et al., 2020). Financial stability allows parents to invest in language development through private tutoring and immersion experiences that cultivate oral fluency and critical reading abilities. Moreover, elevated income levels are associated with reduced household stress, enabling parents to engage more consistently in academic support activities that strengthen students' motivation and study habits (Sorhagen, 2013). The interplay between family income and school engagement further compounds its predictive capacity, as financially secure families are better positioned to foster school-home collaboration, which reinforces academic accountability and achievement (Crosnoe, 2009). These mechanisms collectively demonstrate that annual family income exerts a multifaceted influence on English performance by shaping both cognitive development and the broader educational context. Annual family income is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Parental involvement represents a predictor of secondary students' English performance by shaping cognitive, affective and environmental conditions that facilitate language learning. Family engagement enhances students' academic motivation, fosters positive attitudes towards learning and provides consistent linguistic stimulation that supports English language acquisition (He and Thompson, 2022). Empirical findings indicate that parental involvement exerts both direct and indirect effects on English achievement through its influence on learning engagement and self-regulated study behaviours (Wang et al., 2023). Moreover, parental support amplifies students' access to educational resources, scaffolds home literacy practices and reinforces academic persistence, all of which contribute to vocabulary growth and reading comprehension crucial for English proficiency (Niehaus and Adelson, 2014). These cumulative influences underscore the multifaceted role of parental involvement in creating enriched learning contexts that promote sustained academic success in English development. Parental involvement in student's learning is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Past English performance serves as a robust predictor of current English performance by capturing stable linguistic competencies and learning trajectories. Prior success reflects established vocabulary breadth, syntactic proficiency and reading comprehension skills that directly transfer to subsequent language tasks (Musa, 2024). Longitudinal assessments consistently demonstrate that earlier English proficiency accounts for significant variance in future academic outcomes, suggesting cumulative reinforcement of foundational language skills (Brown, 2024 September). Positive feedback loops, whereby prior achievement enhances self-efficacy and motivation, further contribute to sustained engagement in English learning (Miao et al., 2025). These mechanisms align with theoretical models of skill consolidation, where early mastery facilitates deeper processing and application of complex linguistic structures over time. Consequently, past English performance encapsulates both cognitive competence and affective engagement, rendering it a highly reliable indicator for predicting ongoing English performance among secondary learners. Past English performance is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Socioeconomic status (SES) constitutes a predictor of secondary students' English performance by shaping cognitive, motivational and environmental learning conditions. Higher SES affords access to enriched home literacy environments, diverse educational resources and quality instructional opportunities that directly support English language acquisition (He and Thompson, 2022). Students from higher SES backgrounds typically engage with extensive vocabulary exposure and complex linguistic interactions, facilitating superior reading comprehension and writing proficiency. Moreover, SES influences parental involvement and educational expectations, both of which mediate students' academic engagement and perseverance in language learning tasks (Wang et al., 2023). Empirical models further confirm that SES significantly predicts academic performance across disciplines, including English, through its interaction with institutional supports and student demographics (Schlendorf et al., 2025). These findings underscore that SES encapsulates multifaceted cognitive, affective and contextual mechanisms, establishing its strong predictive validity for English performance in secondary education. SES is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Teacher's qualification constitutes a predictor of secondary students' English performance by shaping instructional quality and academic outcomes. Highly qualified teachers possess deeper content knowledge, advanced pedagogical skills and effective assessment literacy, enabling them to deliver cognitively demanding English instruction that fosters linguistic proficiency (Booker and Holter, 2025). Empirical evidence indicates that teacher expertise enhances students' engagement in achievement-oriented learning goals, facilitating sustained progress in complex language acquisition tasks (Feng and Li, 2024). Moreover, qualified teachers are better equipped to design curriculum-aligned assessments and provide targeted feedback that reinforces vocabulary development, reading comprehension and writing competence (Brown, 2024 September). Institutional analyses further confirm that systematic variations in teacher qualification levels contribute to disparities in English academic outcomes, underscoring the structural significance of teacher credentialing in language education (Shi, 2025). Collectively, these findings validate that teacher's qualification encapsulates critical cognitive and instructional mechanisms that directly influence English performance. Teacher's qualification is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Teacher's self-efficacy represents a predictor of secondary students' English performance by shaping instructional quality, classroom engagement and academic outcomes. Teachers with high self-efficacy demonstrate stronger instructional adaptability, greater persistence in addressing student needs and more effective use of pedagogical strategies, all of which contribute to enhanced English language acquisition (Rivera and Li, 2025). Elevated teacher self-efficacy fosters positive classroom environments that promote student motivation, participation and sustained cognitive engagement, critical for developing complex linguistic competencies. Moreover, self-efficacious teachers are more likely to adopt student-centred instructional goals and foster individualized feedback, directly supporting reading comprehension, vocabulary acquisition and written fluency (Daumiller et al., 2025). These dynamic teacher behaviours collectively establish cognitively stimulating learning contexts that enable students to consolidate and apply English language skills. Thus, empirical evidence underscores that teacher's self-efficacy encapsulates a constellation of instructional and psychological mechanisms that directly enhance English performance. Teacher's self-efficacy is also the best predictor for the two algorithms: DT, and SVMs.

Teaching method constitutes a predictor of secondary students' English performance by shaping instructional effectiveness and learner engagement. Approaches that promote mastery-oriented classroom structures foster positive expectancy-value beliefs, which directly enhance academic motivation and English language achievement (Zhang et al., 2025). Instructional methods grounded in positive psychology further stimulate emotional engagement and resilience, facilitating sustained language acquisition even in challenging learning contexts (Miao et al., 2025). Moreover, adaptive teaching methodologies that integrate technology-mediated scaffolding personalize instruction, optimize feedback and create enriched learning environments conducive to vocabulary growth, reading comprehension and writing proficiency (Zeng, 2025). These dynamic pedagogical processes enable learners to consolidate complex linguistic competencies and transfer skills across varied English language tasks. Collectively, empirical evidence underscores that teaching method encapsulates cognitive, motivational and technological dimensions that directly contribute to enhanced English performance in secondary education. Teaching method is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

Past total scores serve as a predictor of secondary students' English performance by capturing broad cognitive abilities, academic consistency and cross-domain learning behaviours. High overall academic achievement reflects general literacy, problem-solving competence and metacognitive regulation, all essential for complex language acquisition (Brown, 2024 September). Empirical analyses incorporating multilingual competencies further demonstrate that students with strong cumulative academic profiles exhibit superior language adaptability and transfer skills, facilitating English proficiency development (Ali et al., 2025). Advanced machine learning models validate that aggregate academic indicators, including prior total grades, significantly predict English listening comprehension and reading outcomes by identifying stable performance trajectories (Jiang, 2025). Regression-based studies confirm that cumulative academic metrics encompassing mathematics, English and related subjects offer reliable forecasting power for English achievement due to shared cognitive and motivational constructs (Alnoman, 2024 November). Collectively, these findings underscore that past total scores encapsulate integrated cognitive, affective and behavioural predictors, establishing them as indicators of English performance in secondary education. Past 4 total scores is also the best predictor for the three algorithms: DT, ANNs, and SVMs.

For RQ1c, this study identified the most important predictors of NCEE English performance using the NB algorithm. Annual family income was the strongest factor, supporting access to quality learning resources and environments. Parental involvement enhanced motivation, literacy practices, and self-regulation. Past English performance reflected stable language skills and learning continuity. Socioeconomic status influenced home learning conditions and academic engagement. Teacher's qualification and self-efficacy shaped instructional quality, student participation, and language performance. Teaching method that emphasized motivation, personalization, and emotional engagement improved performance. Finally, past total scores captured general cognitive and behavioral traits, serving as reliable indicators of English performance.

This study connects to Bronfenbrenner's framework by showing how family income and parental involvement (exosystem and microsystem) interact with teacher factors (microsystem) to affect language acquisition. Socioeconomic status and past performance underscore macrosystem influences on educational equity. Limitations include the lack of longitudinal data to track developmental changes and potential confounding variables. For policy, education authorities should allocate funds for low-income families and provide professional development on culturally responsive teaching. Practitioners can enhance parental engagement through workshops and use differentiated instruction to address diverse ecological backgrounds.

5.2.2 The Highest Accuracy Predictive Model to Predict NCEE Performance in Chinese, Mathematics, and English

The confusion matrix is a key tool used to evaluate the performance of predictive models, helping understand the accuracy of predictions for each class. Based on the confusion matrix, researcher derives metrics such as AUC, accuracy, F1-Score, precision, recall, and MCC. These metrics provide a comprehensive view of model performance, assessing its ability to make correct predictions while considering false positives and negatives. In this discussion, researcher will analyse the results from different algorithms. The focus will be on answering RQ4, RQ5, and RQ6, and the results discovered which predictive model has the highest accuracy in predicting Chinese, Mathematics and English performance using NB, DT, ANNs, SVMs algorithms. The upcoming paragraphs discuss the results, compare the effectiveness of each algorithm, and provide insights into their suitability for predicting student performance across different subjects.

For RQ2a, the comparative analysis of predictive models demonstrates that the NB algorithm yields superior accuracy in forecasting students' Chinese performance, achieving 95.1%, surpassing ANNs, DT and SVMs. The NB model's probabilistic framework effectively manages categorical data distributions and inter-class separations, contributing to its higher precision and recall. In contrast, while ANNs approached NB in performance, its slightly lower precision indicates sensitivity to misclassification, particularly among high performers. Both DT and SVMs exhibited comparable but lower accuracies, reflecting limitations in handling multi-class imbalances inherent in language performance datasets. Thus, NB emerges as the most effective algorithm for this prediction task.

For RQ2b, the comparative evaluation reveals that the NB algorithm demonstrates superior predictive accuracy in Mathematics performance, achieving 96.4%, surpassing ANNs, SVMs and DT. The probabilistic nature of NB effectively captures conditional dependencies within the academic data, enhancing its classification precision and recall. ANNs and SVMs follow closely, exhibiting high but slightly lower accuracies, reflecting sensitivity to class overlaps and misclassifications, particularly in high-performing groups. Decision Trees exhibit the lowest accuracy, suggesting limitations in managing complex multi-class distributions. These findings underscore NB's robustness in handling categorical academic variables and its suitability for predicting Mathematics performance in secondary education contexts.

For RQ2c, The comparative analysis indicates that the Naive Bayes (NB) algorithm demonstrates superior performance in predicting English performance, attaining 90.7% accuracy, surpassing SVMs, ANNs and DT. The probabilistic structure of NB effectively accommodates the conditional dependencies and categorical nature of language performance data, enhancing its classification stability. While SVMs approach NB in accuracy, their reduced precision reflects sensitivity to class boundary overlaps, particularly in high-performing groups. ANNs and DT show comparatively lower accuracies, suggesting limitations in handling nuanced linguistic variations. These findings affirm the robustness of NB in capturing the multidimensional characteristics essential for predicting English proficiency in secondary education.

The consistent superiority of the NB algorithm across the prediction tasks of Chinese, Mathematics and English performance reveals its robust applicability within academic performance modelling. NB's probabilistic foundation allows it to accommodate the conditional dependencies between multiple academic indicators while managing class imbalances inherent in educational datasets (Zheng & Na, 2021). In predicting English proficiency, emotional variables and linguistic features interplay with cognitive variables, generating complex feature spaces that NB effectively simplifies through its independence assumptions (Huynh-Cam et al., 2022). Empirical evaluations confirm that NB achieves superior classification precision in English learning contexts compared to more computationally intensive algorithms (Zheng & Li, 2024). Furthermore, the algorithm's capacity to integrate affective variables has been validated in cross-national education datasets, enhancing early detection of struggling language learners (Ma et al., 2025). This capacity is particularly valuable in multilingual education systems, where overlapping language proficiency and cognitive domains challenge the discriminatory power of alternative algorithms (Wang et al., 2022). The NB model's simplicity supports effective generalization even with limited sample sizes, which is frequently observed in stratified educational assessments (Batool et al., 2023).

In Mathematics performance prediction, NB continues to exhibit superior performance by effectively capturing discrete categorical patterns often embedded within mathematical achievement distributions. Mathematical competence is frequently influenced by interdependent skill clusters involving problem-solving strategies, procedural fluency and conceptual understanding, all of which produce highly structured but correlated features (Liu et al., 2022). The NB classifier's ability to model these dependencies with probabilistic reasoning allows it to outperform tree-based models and neural networks, which may overfit local patterns without leveraging global data distributions (Roslan & Chen, 2023). Comparative studies have demonstrated that NB achieves stable predictive accuracy even when feature sets expand to include metacognitive factors, learning behaviours and prior subject mastery, while complex models often suffer from reduced generalizability (Zhang et al., 2021). Furthermore, NB's resilience to multicollinearity, which frequently arises in education datasets, enhances its classification reliability in mathematical domains where feature interdependencies are pronounced (Zhang et al., 2022). These findings align with broader observations that educational data mining benefits from models capable of leveraging discrete probabilistic relationships rather than solely optimizing hyperparameter complexity (Ali et al., 2025). The cumulative evidence supports NB's consistent effectiveness in predicting mathematics outcomes across diverse student populations.

The superior predictive capability of NB extends to Chinese performance modelling, where language mastery involves hierarchical constructs combining vocabulary acquisition, syntactic development and semantic processing. Chinese language proficiency prediction presents unique challenges due to its logographic nature and complex character structures, which interact with cognitive and metalinguistic competencies (Zhang et al., 2021). NB demonstrates high sensitivity in managing these multifactorial dependencies by simplifying multidimensional linguistic variables into conditional probabilities (Batool et al., 2023). Comparative investigations reveal that NB consistently achieves higher classification accuracy than support vector machines, decision trees and artificial neural networks in predicting students' comprehensive language proficiency across Chinese cohorts (Zheng & Na, 2021). Moreover, NB's probabilistic handling of categorical features enables robust modelling of student heterogeneity across diverse instructional contexts, effectively accommodating variations in socioeconomic background, emotional engagement and prior academic achievement (Wang et al., 2022). The empirical synthesis confirms that the NB model's simplicity does not compromise its predictive power but rather enhances stability and interpretability across multiple academic domains (Huynh-Cam et al., 2022; Zheng & Li, 2024). This cumulative evidence underscores the versatility and robustness of NB as a superior classifier for comprehensive academic performance prediction.

5.2.3 The Relationships between Chinese, Mathematics, and English Performance

The results clearly show significant relationships between student performance in Chinese, Mathematics, and English, addressing RQ3.

The predictor rankings suggest subject-specific interdependencies. In Section 4.5.1, English and Mathematics both hold mid and weak levels of importance, respectively, in predicting Chinese performance, indicating an association likely driven by shared cognitive or academic behaviors. However, Section 4.5.2 shows that Chinese and English have minimal predictive value for Mathematics, implying that Mathematics relies on distinct cognitive skills, such as logical reasoning, rather than linguistic abilities. Section 4.5.3 reinforces the link between English and Chinese, with Chinese consistently ranking as a relevant, though not dominant, predictor of English performance. This bidirectional association between the two language subjects may reflect overlapping literacy skills and learning strategies. In contrast, the limited predictive role of language subjects for Mathematics suggests compartmentalized academic domains. These insights highlight the need for subject-specific pedagogical approaches, with integrated support for language development and separate strategies for enhancing mathematical thinking.

The consistently weak correlations between Chinese and Mathematics performance observed across five assessments (ranging from .124 to .267) suggest that mastery in the Chinese language contributes minimally to mathematical achievement. A meta-analysis of 34 studies with nearly 59 000 participants revealed that although overall Chinese language ability relates moderately to mathematics (r = .36), specific linguistic components such as rapid automatic naming exhibit notably weaker associations with calculation skills (Lu et al., 2022). Similarly, an investigation of reading comprehension and mathematics achievement among 666 high school students found that while comprehension skills predict some variance in mathematics scores, the strength of this relationship remains limited (Imam et al., 2013). Further, evidence from a study of minority students indicates that deficits in Chinese language proficiency underlie only a portion of lower mathematics outcomes, underscoring that broader cognitive and instructional factors predominate (Cui et al., 2022). These findings align with current results, in which Chinese performance accounted for little of the variability in Mathematics performance at any measurement point.

In contrast, the moderate and stable correlations between Chinese and English performance (r values from .308 to .416) reflect substantial overlap in the linguistic skills required for both subjects. Cross-linguistic transfer research demonstrates that phonological awareness, decoding, and vocabulary knowledge in the first language facilitate similar competencies in a second language, yielding correlation coefficients in the moderate range (Yang et al., 2017). A recent study of bilingual motivation and engagement among rural Chinese adolescents further confirmed that strong first-language skills bolster second-language learning outcomes through shared metalinguistic awareness (Bai et al., 2025). Moreover, parenting practices linked to socioeconomic status mediate the acquisition of both Chinese and English proficiencies, amplifying their interdependence (Xu & Jin, 2024). These convergent findings substantiate current observation of a meaningful and enduring positive relationship between Chinese and English performance across assessments.

The evolution of Mathematics and English correlations from weak in the first assessment (r = .220) to consistently moderate thereafter (r ≈ .365 - .402) highlights how initial quantitative-linguistic interplay may strengthen as students develop domain-specific skills. Research on English language proficiency in mathematical contexts shows that enhanced English comprehension directly improves the ability to solve word problems, producing moderate correlation coefficients similar to those we observed (Mallika & Mohammed., 2024). An earlier study of bilingual university students found that English proficiency accounts for a significant portion of variance in Mathematics performance, with moderate effect sizes (Yushau & Omar, 2015). Longitudinal data also indicate that improvements in English language skills over time correspond with gains in mathematical reasoning and problem-solving, reinforcing a reciprocal reinforcement mechanism (Henry et al., 2014). Collectively, this literature supports current finding that initial weak quantitative linguistic links can mature into moderate associations as instructional exposure and cognitive integration deepen.

The findings reveal distinct intersubject dynamics shaped by cognitive domains and developmental trajectories. Chinese and English demonstrate a stable, moderate association, likely driven by shared linguistic skills such as vocabulary knowledge, decoding, and metalinguistic awareness. This relationship is reinforced by consistent predictor rankings and moderate correlation coefficients, suggesting overlapping learning mechanisms and mutual reinforcement, particularly in literacy-based competencies. In contrast, Chinese shows minimal predictive power for Mathematics, and their weak correlations suggest largely independent cognitive demands, with logical reasoning outweighing language proficiency. However, the link between Mathematics and English appears more dynamic. While initially weak, the correlation strengthens over time, indicating that as students advance, linguistic proficiency, particularly in reading comprehension plays a greater role in mathematical problem-solving. This evolving association suggests a cumulative effect of language skills on Mathematics performance in word-based contexts. These patterns underscore the importance of subject-specific instructional strategies while also recognizing opportunities for cross-disciplinary skill transfer, especially between language and Mathematics in later stages of learning.

5.2.4 The DT Rules for High Performance in Physics, Chemistry, Biology, Politics, History, and Geography

RQ4a-RQ4f utilise DT rules to match students' high performance in subjects through their different characteristics, which in turn provides scientific recommendations for their subject options. DT algorithm is a simple and powerful tool for classification. It works well with both numerical and categorical data. DT is easy to use and explain because it shows the decision process clearly. It is fast and does not require much data preparation. DT often gives high accuracy because it identifies important patterns in the data. It also avoids overfitting by using pruning techniques. These features make DT a strong choice for many classification tasks (Charbuty & Abdulazeez, 2021; Priyanka & Kumar, 2020; Patel & Prajapati, 2018).

For RQ4a, this study explored the characteristics affecting high performance in Physics during NCEE. These factors include student's self-efficacy and peer relationship.

The findings suggest that self-efficacy plays a critical role in determining high performance in Physics examinations. Students with higher self-efficacy were significantly more likely to excel, a pattern supported by several studies. Self-efficacy enhances both motivation and academic performance, particularly in Physics students (Bottomley et al., 2022). Sawtelle et al. (2012) also demonstrated that higher self-efficacy predicts both retention and success in physics. Furthermore, self-efficacy directly affects academic performance, reinforcing the positive correlation between confidence and performance in Physics (Appiah-Twumasi, 2024). The results show that strategies to improve student's self-efficacy could significantly boost academic success.

Students with high level peer relationship are more likely to sustain high performance, a trend widely supported in recent literature. High peer interaction supports learning through discussion and shared understanding. Students who discuss Physics concepts with classmates often score higher in final exams (Simpfendoerfer et al., 2024). Peer mentoring is also linked to better academic outcomes. Working with a supportive peer helps students grasp difficult ideas and stay motivated (Tian et al., 2025). Peer tutoring further boosts achievement. Students who tutor or are tutored by peers show greater gains in Physics test scores (Luminoque, 2022). This is because peer explanations often match the learner's level of understanding. Sustained collaboration with friends leads to even stronger benefits. When students work in trusted peer groups, they engage more deeply and persist longer in Physics tasks (Pulgar et al., 2022). Over time, such friendships foster higher achievement and more positive attitudes toward Physics. These findings collectively affirm that peer relationship remains a strong indicator of future performance in Physics subject.

For RQ4b, this study identified several key characteristics affecting high performance in Chemistry during NCEE. It highlights several key factors, such as student's self-efficacy, past Biology performance, teacher's qualification, and social support.

The result shows that self-efficacy is important for doing well in Chemistry. This aligns with the findings of Villafañe et al. (2016), who found that students with higher self-efficacy performed better in organic Chemistry. They showed that self-efficacy and performance affect each other over time (Villafañe et al., 2016). This finding is consistent with the work of Avargil (2019), who showed that self-efficacy improved when students learned Chemistry with real-world examples and teacher support. Students with stronger self-efficacy understood Chemistry better and performed well (Avargil, 2019). Alci (2015) also found similar results. His study showed that both self-efficacy and motivation help students perform well in Chemistry. Students with high confidence in their abilities and strong interests often scored better (Alci, 2015). These studies suggest that teachers should focus on building student confidence. When students feel more capable, they may try harder and stay motivated, leading to better results.

The result shows that students' past academic performance can predict future success. Students with solid Biology knowledge often learn Chemistry more easily. Biology gives context for many chemical ideas. A study found that students with stronger past Biology work scored higher in an introductory chemistry-in-biology intervention (Mendez et al., 2024), showing cross-subject gains. Another research project used asynchronous Chemistry modules embedded in Biology courses. It showed that students with better Biology class results also gained more from Chemistry content (Meaders et al., 2025). A meta-analysis found that web-based Biology learning improved student performance in STEM broadly. Improved Biology scores were associated with better achievement in subsequent Chemistry tasks (Vekli & Çalik, 2023). Finally, a study of biochemistry labs reported that students who had done well in Biology were more successful in lab Chemistry exercises. Strong Biology background reduced common mistakes in chemistry-based experiments (Costabile et al., 2024). These findings together show that good Biology performance helps Chemistry learning. Biology provides foundation for chemical terms, lab habits, and conceptual links between molecules and processes. When students enter Chemistry with firm Biology skills, they grasp topics faster. Strong Biology thus acts as a scaffolding for Chemistry success.

Teachers with high qualifications often help students perform better in Chemistry. For example, in a recent study, teachers with Chemistry degrees and more experience led to significantly higher student scores (Cleopas & Onwuchekwa, 2024). This shows that qualification directly supports student achievement. Similarly, a study in Nigerian colleges found that students taught by highly qualified Chemistry teachers performed better in final exams (Apampa et al., 2024). In addition, another study in Bayelsa State confirmed that students with qualified and experienced Chemistry teachers scored higher in national tests (Nnoli, 2024). This supports the idea that teacher credentials are key to student success. Moreover, findings from Ethiopia showed that students taught by more qualified teachers earned higher scores across science subjects (Engida, 2024 July). Thus, teacher's qualification influences science learning more broadly. Overall, teachers with higher qualifications bring deeper subject knowledge and better teaching strategies. They explain Chemistry more clearly, guide lab work better and help students learn effectively. Therefore, improving teacher's qualification can greatly benefit Chemistry education.

Students who receive social support tend to improve in Chemistry. First, a study with 1,260 Chinese high school students found that teacher and peer support promoted Chemistry motivation and self-efficacy (Huangfu et al., 2023). This support led to better engagement in Chemistry lessons. A recent article showed that targeted academic and peer support in general Chemistry increased engagement and scores for underrepresented students (Kumari et al., 2025). Thus, structured social help can improve performance in Chemistry classes. Finally, another study reported that learning assistants and instructors who offered active support helped students feel more confident and achieve more in general Chemistry (Donis et al., 2023). Their guidance improved lab skills and course outcomes. Together, these studies show a clear link. When students receive help from teachers, peers, or structured support programs, they gain confidence. As a result, they study more, participate better, and score higher in Chemistry.

For RQ4c, this study identified key characteristics affecting high performance in Biology during NCEE based on DT rules. It is primarily influenced by three key factors, such as total scores in final-term examination before subject options in 2023, student's self-efficacy, and social support.

Students who perform well across many subjects also often get good scores in Biology. For example, one study found that incoming college GPA was the strongest predictor of course scores in large life science classes (Creech & Sweeder, 2012). Another research showed that high school GPA is the best predictor of first-year science degree success, and it explained about 20% of the variance in biology-specific performance (Dorta - Guerra et al., 2019). Also, data from a private university revealed that verbal SAT and high school GPA significantly predicted first-semester Biology exam scores (Montague, 1995). Furthermore, another study found that prior knowledge when combined with high school GPA strongly predicts first-year Biology success (Binder et al., 2019). These studies support the idea. In short, general academic strength and prior preparation lead to solid performance in Biology too. Therefore, to enhance Biology performance, schools should focus on boosting overall GPA, relevant prior knowledge, and strong academic habits early on.

Student’s self-efficacy significantly influences future Biology performance. This aligns with Musgrove et al. (2024), who found that self-efficacy contributes to students’ persistence and achievement in Biology education. Udeh-Aloysius and Achufusi (2024) reported a positive correlation between self-efficacy and academic achievement in Biology among secondary school students which is consistent with this finding. Furthermore, Ewere (2023) demonstrated that self-efficacy is a significant predictor of Biology students’ academic performance, as it helps build confidence and persistence in learning tasks. Therefore, students with higher self-efficacy are more likely to achieve better performance in Biology subject.

Students with strong social support often record better scores in biology. For instance, one study of Chinese middle schoolers found that social support directly improved academic results through greater self-efficacy and engagement (Zhang & Qian, 2024). Support from family, supervisors, and institutions was shown to boost postgraduate students’ academic outcomes via positive academic emotions (Zhang & Ren, 2024). Moreover, biology-specific research revealed that belonging to a learning community raised grades, confidence, and persistence in biology classes (Wu et al., 2024). Another study of special-needs students in Iraq reported that family and friend support predicted higher academic achievement overall (Saeed et al., 2023). These studies confirm that social backing benefits Biology students too. Therefore, when schools ensure support systems, Biology grades rise. Thus, programs should focus on emotional, academic, and community support. In doing so, they can lift Biology performance and foster student success across science subjects.

For RQ4d, the results highlighted several critical factors affecting students’ high performance in History using DT rules. The primary characteristics are social support, total scores in admission examination in 2022, socioeconomic status, and past Politics performance. These factors collectively affect the likelihood of achieving high NCEE History performance.

Receiving support from family, teachers, and friends helps students feel more confident. This boosts their belief in their own skills and encourages them to try harder in class (Zhang & Qian, 2024). In fact, social support directly improves test scores for Chinese middle schoolers through stronger engagement in learning (Zhang & Qian, 2024). Besides, peer relationships alone are a clear predictor of better grades. When students feel supported by their classmates, their motivation and class focus go up. This leads to higher scores (Shao et al., 2024). Moreover, teacher support is especially strong. Students who feel cared for in class are more likely to focus on lessons and remember important facts (Martinot et al., 2022). Likewise, a broad study across multiple countries found that early adolescents who reported higher support from parents, peers, and teachers scored higher in academic scores (Ahmed et al., 2010). Hence, middle schoolers who get help at home, friendship from classmates, and encouragement from teachers learn more History content. They study more, stay calm under exam stress, and enjoy class better. Although the above studies did not specifically mention the History subject, it is possible to infer from them the significant influence of social support on various subjects, including the History subject as well.

Middle school students with strong total scores tend to also excel in History. This relationship is well supported by recent studies. For example, when students develop a strong academic self-concept, their success tends to extend across multiple subjects (Rost & Feng, 2024). This shows that a positive general academic mindset fosters individual subject performance. Teacher stability plays a crucial role in maintaining this pattern. One study found that irregular teacher turnover negatively impacted subject-specific outcomes such as History. However, students with strong general academic performance were better able to withstand these disruptions (Aeschlimann et al., 2019). Thus, prior total scores can act as a buffer. Moreover, the link between self-concept and academic achievement is both subject-specific and influenced by total scores. Students with a high general academic self-concept are more likely to excel in History as well (Marsh, 1992). Notably, these effects are not random but reflect a consistent pattern across subjects. Finally, students' values toward History often grow stronger through prior academic success. In other words, total scores enhances both motivation and outcomes in History (Schneider & Wolff, 2023). Altogether, these findings clearly indicate that excellent total scores increases the likelihood of success in History.

Students from higher socioeconomic status (SES) families often perform better in humanities subjects, such as History. Several studies support this. Higher SES provides families with more educational resources, including books, art, and cultural experiences. These enrich learning in humanities fields (Park & Yi, 2025). Moreover, schools in higher SES areas typically offer better teaching quality and more developed humanities programs, such as in economics and literature (Akanni & Adewole, 2025). This gives students an advantage in both understanding and performance. Family environment also plays a strong role. Parents from higher SES often promote critical thinking and reading, which supports achievement in humanities subjects (Luke & Jude, 2025). Finally, a global study confirms that SES strongly correlates with academic success in various humanities subjects. Students from lower SES backgrounds often lack access to supportive learning environments, which impacts their outcomes (Anderson, 2025). In short, higher SES provides both material and cultural benefits, helping students perform better in humanities education, including History.

Students who perform well in Politics often do well in History too. These two subjects share key skills. Both require critical thinking and understanding of complex social issues. Research shows that credit achievement in American Government (a Politics subject) and US History is strongly correlated. Students who excel in one often succeed in the other (Stark, 2024). Another study found that humanities education promotes skills like argumentation and evaluation (Makvandi et al., 2025). These skills support learning across related subjects, including politics and history. Furthermore, teaching methods such as team-based learning improve engagement and achievement in history classes. This also enhances skills used in political studies (Jebu & Raddam, 2024). Finally, educational data from Sri Lanka shows that academic success in civic education links closely with higher performance in History. This is due to overlapping content and shared literacy demands (Bandula et al., 2024). So, politics and history support each other. Success in one subject often predicts success in the other.

For RQ4e, the DT analysis identified several characteristics contributing to high performance in Politics. These include total scores in mid-term examination before subject options in 2022, and teacher's education level.

Students who do well past scores often do well in Politics too. This is because good study habits and thinking skills help in many subjects. A study showed that students with higher overall achievement also scored higher in civic education (Arnzen, 2024). Strong readers and critical thinkers tend to perform better in politics-related topics. Another study found a clear link between overall grades and civic knowledge (Longo & Perrotti, 2023). When students do well in all classes, they also understand political ideas better. This shows that general academic skills help in civics learning. Research also showed that teaching civics in engaging ways helps students perform better in school overall (Watson et al., 2025). This suggests a two-way relationship: good academic skills help in civics, and good civics teaching boosts general achievement. Finally, a study on learning strategies found that students who gained skills in government and civics classes also improved their overall academic performance (Kekeba et al., 2025). This means that success in school and in politics often go hand in hand. So, strong past total scores supports better performance in Politics studies.

Students perform better in Politics studies when their teachers have higher education levels. A recent study showed that teacher qualifications in civics education strongly impact student achievement in political knowledge (Arnzen, 2024). Well-trained teachers can explain complex topics more clearly. Another study found that teacher education and training in Politics help students develop better civic knowledge and understanding (Meriläinen & Hiljanen, 2024). Teachers with advanced degrees can teach deeper content and skills. Good teacher education also improves how civic values are taught. A study found that well-educated teachers foster better political tolerance and civic outcomes in students (Torres, 2024). This helps students engage more with political topics.

Teacher education affects classroom outcomes in government studies too. A study showed that teachers with higher education levels and training in government teaching lead to better student results (Khan et al., 2025). Students learn more when their teachers are better prepared. In short, higher teacher's education level supports better student Performance in political and civic studies.

For RQ4f, the rule derived from DT analysis identified several key factors affecting students' high performance in Geography: parental involvement in student's learning and past History performance.

Parental involvement is important for student learning. Many studies show that when parents support their child's education, the child performs better. This is also true in Geography. A study showed that secondary school students with more parental support had better knowledge and attitudes in Geography (Wahelo et al., 2025). Parents' education level and participation in school activities help students to succeed in Geography classes. In general, parental involvement improves academic achievement. One study found that parents' support shapes students' academic success (Jung, 2025). When parents help with homework and school projects, students are more motivated to learn. Another study confirmed that parental involvement is a strong predictor of academic success (Basnyat, 2025). Active parental participation in school improves both academic and social outcomes. Further research showed that meaningful parental involvement promotes better student learning (Webb, 2025). Schools that work closely with parents often see higher academic performance in students. Thus, when parents are more involved in their child's education, the child often performs better. This applies to Geography as well as to other school subjects.

Students who perform well in History often do well in Geography too. This is because both subjects use similar skills. Reading, critical thinking, and understanding events over time and space help students in both areas. A recent study found that classroom practices that improve History learning also support Geography achievement (Materu et al., 2025). Good teaching in one subject can help in another. Another study showed that using new tools like social media and images helps students do better in history and Geography (Watson et al., 2025). This shows that when students build skills in History, they can use them in geography too. Teachers also find that mixing History and Geography topics helps students see connections. One project showed that combining lessons on local History and Geography improved student understanding in both subjects (Luke & Jude, 2025 March). In Japan, textbooks now often link History and Geography topics. A study found that this approach helped students improve their performance in both subjects (Kamogashira, 2024). When students learn to see how History shapes places, their Geography learning improves. In short, strong performance in History supports better learning in Geography.

5.3 Chapter Summary

This section provides an in-depth discussion on the following issues: "The Most Important Factors to Predict NCEE Performance in Chinese, Mathematics, and English", "The Highest Accuracy Predictive Model to Predict NCEE Performance in Chinese, Mathematics, and English", "The Relationships between Chinese, Mathematics, and English Performance", and "The Rules for High Performance in Physics, Chemistry, Biology, Politics, History, and Geography". These findings collectively reveal key insights into performance prediction and subject interrelations, offering valuable recommendation for academic planning and decision-making. The next chapter describes the conclusion of this study.

CHAPTER 6

CONCLUSION

6.1 Introduction

This chapter discusses the key findings of the study (see Table 6.1), focusing on the significant implications of the study and offering targeted expectations for future research by addressing its limitations.

Table 6.1: The Key Findings of the Study

Area	Key Findings
Top Performance Predictors	Chinese: motivation and teaching method. Mathematics: test anxiety and parental education level. English: family income and parental involvement
Best Predictive Model	NB achieved highest accuracy for compulsory subjects. DT provided clear rules for subject options
Subject Correlations	Chinese was less related to Mathematics, while English was moderately related to both Chinese and Mathematics
Subject Recommendation	Based on Decision Tree rules using student profiles (e.g., self-efficacy, peer relationship, scores, teacher qualification). 12 specific subject combinations are recommended

6.2 Implications

6.2.1 Performance Prediction of Compulsory Subjects

The discussion reveals that distinct constellations of predictors drive NCEE outcomes in each core subject. For Chinese, intrinsic motivation level commands the highest weight (0.845), followed by teaching method, gender, and prior subject metrics, underscoring the roles of learner drive, pedagogical design, and foundational literacy skills in shaping language performance. In Mathematics, test anxiety emerges as the paramount predictor, with parental education levels, socioeconomic status, peer relationship, self-efficacy, and instructional factors jointly informing student performance, highlighting the interplay between affective regulation, family background, and classroom context. For English, annual family income constitutes the most influential factor (weight 0.605), alongside parental involvement, cumulative English scores, socioeconomic status, teacher qualification and self-efficacy, teaching method, and aggregate achievement metrics, reflecting how economic resources, home support, and instructional quality converge to facilitate language proficiency.

Comparative evaluation of predictive models demonstrates that NB consistently outperforms alternatives across all three subjects. NB achieves 95.1% accuracy for Chinese, 96.4% for Mathematics, and 90.7% for English, surpassing ANNs, SVMs, and decision trees in classification precision, recall, F1-score, and MCC metrics. These outcomes affirm NB's probabilistic framework as especially adept at handling categorical and imbalanced educational data. The strong performance of NB models validates their utility for early identification of at-risk learners, while the robustness of DT and ANNs in English prediction (with 91.3% accuracy) signals the potential of hybrid approaches for capturing nonlinear feature interactions.

Correlational analyses elucidate the interdependencies among subject performances. Chinese and Mathematics exhibit consistently weak correlations ($r = 0.124 - 0.267$), indicating largely distinct cognitive demands, whereas Chinese–English ($r = 0.308 - 0.416$) and Mathematics-English ($r \approx 0.365 - 0.402$) associations fall in the moderate range. These patterns suggest that linguistic competencies transfer reciprocally between Chinese and English, while quantitative reasoning increasingly supports language-based problem solving as students advance. The evolving strength of Mathematics–English links underscores how cumulative domain integration and metacognitive development deepen cross-disciplinary synergies over time.

Grounded in these insights, educators can deploy data-driven, cross-subject predictive models to proactively flag low-scoring students before high-stakes assessments. By monitoring composite risk profiles from Chinese, Mathematics, and English predictors and by leveraging their mutual reinforcement schools can tailor early interventions (e.g., anxiety reduction workshops, family-engagement initiatives, differentiated teaching strategies) to individual needs. Such targeted scaffolding promises to elevate learning trajectories, narrow performance gaps, and ultimately enhance NCEE outcomes for students most in need.

6.2.2 Recommendation of Subject Options

This study confirms that a reliable recommendation framework can be built by matching each student's distinctive profile defined here by self-efficacy, peer relationship, single-subject performance, teacher's qualification, past total scores, social support, parental involvement in student's learning, socioeconomic status, and teacher's education level to DT rules that maximise success in NCEE. The following guidance synthesises all RQ8–RQ13 findings into evidence-based pathways for the twelve official subject options. Every statement links a verified constellation of predictors to the option likeliest to sustain high grades across the three chosen subjects.

Students with high self-efficacy and strong peer relationships are well-suited to prioritize Physics. When student's self-efficacy registers moderate or high, Biology shows at least moderate recent performance, an experienced Chemistry teacher is confirmed, and steady social support exists, Option 1 (Physics, Chemistry, Biology) capitalises on three subjects that each meet their full decision-tree criteria. If Biology performance is only average yet past total scores are moderate or high and the Politics teacher holds an advanced degree, Option 2 (Physics, Chemistry, Politics) maintains laboratory strength while exchanging Biology for Politics, whose rule depends on past total scores and teacher's education level. When verified parental involvement in student's is high and recent History performance are strong factors essential for Geography. Option 3 (Physics, Chemistry, Geography) becomes advisable, provided the original Physics-Chemistry anchors remain intact. Should Chemistry be weakened by limited teacher's qualification or patchy social support, yet

Biology remains strong, and Politics still satisfies its rule, Option 4 (Physics, Biology, Politics) offers a balanced replacement. If Geography predictors outrank Politics alongside a robust Biology anchor, teachers should adopt Option 5 (Physics, Biology, Geography). Option 6 (Physics, Politics, Geography) applies only when Physics stands as the sole science anchor while Politics and Geography each independently meet their DT requirements. Each option aligns strictly with validated predictors, ensuring coherence across the chosen trio.

Students whose profiles meet the History anchor verified social support, moderate or high past total scores, elevated socioeconomic status, and at least moderate Politics performance should choose History first. When Politics and Geography simultaneously fulfil their own predictors, Option 7 (History, Politics, Geography) unites three subjects whose DT conditions are fully satisfied. If Geography is unsupported but Chemistry meets its entire rule set through student's self-efficacy, Biology background, teacher's qualification, and social support, Option 8 (History, Politics, Chemistry) becomes the stronger choice. Where Biology, not Chemistry, satisfies its anchor alongside History and Politics, Option 9 (History, Politics, Biology) maintains the Humanities-Science balance endorsed by the model. When Geography pairs with History and Chemistry is also anchored, Option 10 (History, Geography, Chemistry) is recommended; if Biology replaces Chemistry under the same Geography-History core, Option 11 (History, Geography, Biology) proves more appropriate. A less common profile joins anchored History and Biology with Chemistry while both Politics and Geography lack sufficient social support; here Option 12 (History, Biology, Chemistry) is indicated. Each recommendation follows the exact constellation of validated predictors, avoiding speculative substitutions and conserving examination stability for each student.

6.3 Limitations and Future Research

This study is subject to several notable limitations. The sample is drawn exclusively from a single urban high school (High School F), which restricts the generalizability of the findings. The results may not be representative of the broader student population across China, particularly those in rural schools which often operate under different resource constraints and socio-economic conditions. This introduces a potential for regional bias, as urban institutions typically benefit from superior educational resources and cater to students from higher socio-economic backgrounds, factors that significantly influence academic outcomes. Consequently, the applicability of these findings to disadvantaged or rural educational contexts remains unverified. Methodologically, the study is constrained by its reliance on only four predictive algorithms: Naïve Bayes, Decision Trees, Artificial Neural Networks, and Support Vector Machines. This scope excludes other potent and contemporary techniques, notably ensemble methods like Random Forest, which could potentially offer different insights or superior predictive performance.

To address these limitations, future research should prioritize expanding the scope, methodological rigor of the inquiry, and implement longitudinal follow up. A critical next step involves validating the current findings by replicating the study across a more diverse array of educational institutions. This should include rural high schools, vocational schools, and institutions from various socio-economic strata to test the robustness and universality of the predictive models. Such multi-site studies would help quantify and mitigate the regional bias identified in this work. Furthermore, subsequent investigations should significantly broaden the methodological approach by incorporating and comparing a wider range of machine learning algorithms. A primary focus should be on implementing and evaluating ensemble methods, such as Random Forest and Gradient Boosting, which are known for their high accuracy in complex prediction tasks. Extending the analysis to include deep learning architectures or time-series models could also uncover more nuanced patterns in student performance data, ultimately leading to more powerful, generalizable, and accurate EDM frameworks.

REFERENCES

Abbasian, R., Hadian, B., & Vaez-Dalili, M. (2020). Examination of the role of family socio-economic status and parental education in predicting English as a foreign language learners' receptive skills performance. Cogent Education, 7(1), 1710989.

Abdelkader, H. E., Gad, A. G., Abohany, A. A., & Sorour, S. E. (2022). An efficient data mining technique for assessing satisfaction level with online learning for higher education students during the COVID-19. IEEE Access, 10, 6286-6303.

Abid, A., Jan, A., Khan, I. U., Zeb, A., & Ahmad, M. (2021). Role of parental socioeconomic status on their children's education at District Mardan of Khyber Pakhtunkhwa, Pakistan. Geografia, 17(2), 93-101.

Abrori, M. A. M., Syukur, A., Affandy, A., & Soeleman, M. A. (2022). Improving C4. 5 Algorithm Accuracy With Adaptive Boosting Method For Predicting Students in Obtaining Education Funding. Journal of Development Research, 6(2), 137-140.

Abuhassna, H., Al-Rahmi, W. M., Yahya, N., Zakaria, M. A. Z. M., Kosnin, A. B. M., & Darwish, M. (2020). Development of a new model on utilizing online learning platforms to improve students' academic performances and satisfaction. International Journal of Educational Technology in Higher Education, 17, 1-23.

Acito, F. (2023). Introduction to KNIME. In Predictive Analytics with KNIME: Analytics for Citizen Data Scientists (pp. 21-52). Cham: Springer Nature Switzerland.

Acosta-Gonzaga, E., & Ramirez-Arellano, A. (2021). The influence of motivation, emotions, cognition, and metacognition on students' learning performance: A comparative study in higher education in blended and traditional contexts. Sage Open, 11(2), 21582440211027561.

Adams, R. V., & Blair, E. (2019). Impact of time management behaviours on undergraduate engineering students' performance. Sage Open, 9(1), 2158244018824506.

Adi, A. C., Lestari, D. P., Elsa, F. S. S., & Sabui, Y. (2022). Online School Sentiment Analysis in Indonesia on Twitter Using The Naïve Bayes Classifier and Rapid Miner Tools. Int. J. Innov. Sci. Res. Technol, 71(1), 785-788.

Aeschlimann, B., Herzog, W., & Sander, F. (2019). Irregular Teacher Turnover and Student Academic Achievement in High Schools: A Study in the Subjects Mathematics, German, French and History. Journal of Education and Learning, 8(2), 25-36.

Agasisti, T., Avvisati, F., Borgonovi, F., & Longobardi, S. (2018). Academic Resilience: What Schools and Countries Do to Help Disadvantaged Students Succeed in PISA. OECD Education Working Papers, No. 167.

Ahmad, N. (2023). The Framework of Inclusive Education for Vice Chancellors of Universities Based on the Goals of Sustainable Development through Visual Digital Dashboard. E-Leader International Journal, 16 (1).

Ahmed, A. M., Rizaner, A., & Ulusoy, A. H. (2018). A novel decision tree classification based on post-pruning with Bayes minimum risk. Plos one, 13(4), e0194168.

Ahmed, W., Minnaert, A., van der Werf, G., & Kuyper, H. (2010). Perceived social support and early adolescents' achievement: The mediational roles of motivational beliefs and emotions. Journal of youth and adolescence, 39, 36-46.

Ajibade, S. S. M., Dayupay, J., Ngo-Hoang, D. L., Oyebode, O. J., & Sasan, J. M. (2022). Utilization of Ensemble Techniques for Prediction of the Academic Performance of Students. Journal of Optoelectronics Laser, 41(6), 48-54.

Akanni, O. O., & Adewole, K. A. (2025). Diagnostic and Formative Assessment Strategies on Improving Economics Achievement Among Secondary School Students in Lagos State. NIU Journal of Humanities, 10(1), 95-105.

Akello, G. R. (2020). Influence Of Parents' Involvement In Education On Their Children'S Performance At The Kenya Certificate Of Primary Education In Nyakach Sub-County, Kisumu, Kenya (Doctoral dissertation, University of Nairobi).

Akendita, P. A., Arthur, Y. D., & Asare, B. (2024). Moderating effect of teacher efficacy on the relationship between students' perception of mathematics and students' mathematics achievement. International Journal of Didactical Studies, 6(1), 28932.

Akhrif, O., Benfaress, C., EL Jai, M., El Bouzekri El Idrissi, Y., & Hmina, N. (2022). Completeness based classification algorithm: a novel approach for educational semantic data completeness assessment. Interactive Technology and Smart Education, 19(1), 87-111.

Akinsola, E. F., & Nwajei, A. D. (2013). Test Anxiety, Depression and Academic Performance: Assessment and Management Using Relaxation and Cognitive Restructuring Techniques. Psychology, 4(6A1), 18-24.

Al Shannag, Q. M., Tairab, H. H., Dodeen, H., & Abdel-Fattah, F. (2013). Linking teachers' quality and student performance in the kingdom of saudi arabia and singapore: The impact of teachers' background variables on student performance. Journal of Baltic Science Education, 12(5), 652-665.

Alabdulwahab, S., & Moon, B. (2020). Feature selection methods simultaneously improve the detection accuracy and model building time of machine learning classifiers. Symmetry, 12(9), 1424-1443.

Al-Abyadh, M. H. A., & Abdel Azeem, H. A. H. (2022). Academic performance: influences of university students' self-management and perceived self-efficacy. Journal of Intelligence, 10(3), 55.

Alalawi, S. J. S., Shaharanee, I. N. M., & Jamil, J. M. (2023). Clustering student performance data using k-means algorithms. Journal of Computational Innovation and Analytics (JCIA), 2(1), 41-55.

Alam, A., & Mohanty, A. (2022, November). Business models, business strategies, and innovations in EdTech companies: integration of learning analytics and artificial intelligence in higher education. In 2022 IEEE 6th Conference on Information and Communication Technology (CICT) (pp. 1-6). IEEE.

Alboaneen, D., Almelihi, M., Alsubaie, R., Alghamdi, R., Alshehri, L., & Alharthi, R. (2022). Development of a web-based prediction system for students' academic performance. Data, 7(2), 21-39.

Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Literature Review of Student's Performance Prediction Using Machine Learning Techniques. Education Sciences, 11(9), 552-578.

Alci, B. (2015). The influence of self-efficacy and motivational factors on academic performance in general chemistry course: A modelling study. Educational research and reviews, 10(4), 453-461.

Alemany-Arrebola, I., Rojas-Ruiz, G., Granda-Vera, J., & Mingorance-Estrada, Á. C. (2020). Influence of COVID-19 on the perception of academic self-efficacy, state anxiety, and trait anxiety in college students. Frontiers in psychology, 11, 570017.

Algarni, A. (2016). Data mining in education. International Journal of Advanced Computer Science and Applications, 7(6).

Alghamdi, A. S., & Rahman, A. (2023). Data mining approach to predict success of secondary school students: A Saudi Arabian case study. Education Sciences, 13(3), 293.

Alhadabi, A., & Karpinski, A. C. (2020). Grit, self-efficacy, performance orientation goals, and academic performance in University students. International Journal of Adolescence and Youth, 25(1), 519-535.

Ali, J. A., Abdi, M. K., Ali, T. A., Muse, A. H., Omar, M. A., & Cumar, M. A. (2025). Multilingual competency and academic performance: a machine learning-based analysis of the 2022/2023 Somaliland national primary exam data. Discover Data, 3(1), 1-13.

Alija, S., Beqiri, E., Gaafar, A. S., & Hamoud, A. K. (2023). Predicting Students Performance Using Supervised Machine Learning Based on Imbalanced Dataset and Wrapper Feature Selection. Informatica, 47(1).

Al-Kindi, I., & Al-Khanjari, Z. (2022). A Comparative Study of Classification Algorithms of Moodle Course Logfile using Weka Tool. International Journal for Computers & Their Applications, 29(3).

Almaiah, M. A., Al-Khasawneh, A., & Althunibat, A. (2020). Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic. Education and information technologies, 25, 5261-5280.

Almarabeh, H. (2017). Analysis of Students' Performance by Using Different Data Mining Classifiers. International Journal of Modern Education and Computer Science, 9(8), 9-15.

Alnoman, A. (2024, November). Regression-Based Student Performance Prediction Using the Grades of English, Math, and Programming Courses. In 2024 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD) (pp. 1-6). IEEE.

Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. A. (2022). Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. Computational intelligence and neuroscience, 2022, 4151487-4151487.

Alshareef, F., Alhakami, H., Alsubait, T., & Baz, A. (2020). Educational data mining applications and techniques. International Journal of Advanced Computer Science and Applications, 11(4),729-734.

Alsharif, M., & Rawat, D. B. (2021). Study of machine learning for cloud assisted iot security as a service. Sensors, 21(4), 1034-1045.

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. International Journal of Educational Technology in Higher Education, 17, 1-21.

Anchen, C., & Ying, Z. (2022). Mathematics learning performance: its correlation with chemistry learning performance. Indonesian Journal of Science and Mathematics Education, 5(2), 134-146.

Anderson, I. F. (2025). Pedagogical Innovation and Success Stories in Technical Education: A Case Study at the Technical Education School Nº 2 "Independencia" in Concordia, Entre Ríos, Argentina, and its articulation with national and international publications in Exact, Natural, Social and Technological Sciences. OSF Preprints.

Angrave, L., Jensen, K., Zhang, Z., Mahipal, C., Mussulman, D., Schmitz, C. D., Baird, R. T., Liu, H., Sui, R., Wu, M. S., Kooper, R. (2020, June). Improving student accessibility, equity, course performance, and lab skills: How introduction of ClassTranscribe is changing engineering education at the University of Illinois. In ASEE Annual Conference and Exposition, Conference Proceedings (Vol. 2020, p. 839).

Annuš, N., Takáč, O., Štempeľová, I., & Dancsa, D. (2023). Z and alpha generation teaching methods: digitalization of learning material. International Journal of Advanced Natural Sciences and Engineering Researches, 7(4), 224-229.

Ansari, J. A. N., & Khan, N. A. (2020). Exploring the role of social media in collaborative learning the new domain of learning. Smart Learning Environments, 7(1), 1-16.

Apampa, S. A., Isah, A., & Abubakar, S. (2024). Impact of Teachers' Qualification and Years of Teaching Experience on Academic Performance of Chemistry Students in Colleges of Education, Kaduna State Nigeria. ATBU Journal of Science, Technology and Education, 12(3), 132-140.

Appiah-Twumasi, E. (2024). Predicting Physics Students' Academic Performance: The Impact of Attitude, Self-Efficacy, and Personality Traits. Journal of Research in Education and Pedagogy, 1(2), 107-117.

Araka, E., Oboko, R., Maina, E., & Gitonga, R. (2022). Using educational data mining techniques to identify profiles in self-regulated learning: an empirical evaluation. The International Review of Research in Open and Distributed Learning, 23(1), 131-162.

Arashpour, M., Golafshani, E. M., Parthiban, R., Lamborn, J., Kashani, A., Li, H., & Farzanehfar, P. (2023). Predicting individual learning performance using machine‐learning hybridized with the teaching‐learning‐based optimization. Computer Applications in Engineering Education, 31(1), 83-99.

Arcinas, M. M., Sajja, G. S., Asif, S., Gour, S., Okoronkwo, E., & Naved, M. (2021). Role of data mining in education for improving students performance for social change. Turkish Journal of Physiotherapy and Rehabilitation, 32(3), 6519-6526.

Arens, A. K., Frenzel, A. C., & Goetz, T. (2022). Self-concept and self-efficacy in math: Longitudinal interrelations and reciprocal linkages with achievement. The Journal of Experimental Education, 90(3), 615-633.

Arif, I. B. M., & Rohani, R. B. (2022). Analysis of Industry and Institutional Feedback for Students with Hearing Problems at Politeknik Ibrahim Sultan Using the Weka Application. Education and Tvet Studies, 169-175.

Ariyanto, R. A., & Chamidah, N. (2021, February). Sentiment analysis for zoning system Admission policy using support vector machine and naive bayes methods. In Journal of Physics: Conference Series (Vol. 1776, No. 1, p. 012058). IOP Publishing.

Arnzen, C. J. (2024). Education For Democracy: Three Essays Exploring the Relationship Between Education and Voting (Order No. 31149064). Available from ProQuest Dissertations & Theses Global. (3072904798).

Arslan, Ç. İ. Ğ. D. E. M. (2020). Examining The Relationship Between 5-8th Grade Students's Test Anxiety and Mathematics Anxiety. Acta Didactica Napocensia, 13(1).

Asad, R., Altaf, S., Ahmad, S., Mahmoud, H. A., Huda, S., & Iqbal, S. (2023). Machine Learning-Based Hybrid Ensemble Model Achieving Precision Education for Online Education Amid the Lockdown Period of COVID-19 Pandemic in Pakistan. Sustainability, 15(6), 5431.

Ashish, L., & Anitha, G. (2024, August). Transforming Education: The Impact Of ICT And Data Mining On Student Outcomes. In 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT) (Vol. 1, pp. 675-683). IEEE.

Assem, H. D., Nartey, L., Appiah, E., & Aidoo, J. K. (2023). A Review of Students' Academic Performance in Physics: Attitude, Instructional Methods, Misconceptions and Teachers Qualification. European Journal of Education and Pedagogy, 4(1), 84-92.

Atlay, C., Tieben, N., Hillmert, S., & Fauth, B. (2019). Instructional quality and performance inequality: How effective is teaching in closing the social performance gap? Learning and Instruction, 63, 101211.

Avargil, S. (2019). Learning chemistry: Self-efficacy, chemical understanding, and graphing skills. Journal of Science Education and Technology, 28(4), 285-298.

Baashar, Y., Alkawsi, G., Ali, N. A., Alhussian, H., & Bahbouh, H. T. (2021, July). Predicting student's performance using machine learning methods: A systematic literature review. In 2021 International Conference on Computer & Information Sciences (ICCOINS) (pp. 357-362). IEEE.

Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., Alsariera, Y. A., Ali, A. Q., Hashim, W., & Tiong, S. K. (2022). Toward predicting student's academic performance using artificial neural networks (ANNs). Applied Sciences, 12(3), 1289.

Baber, H. (2020). Determinants of students' perceived learning outcome and satisfaction in online learning during the pandemic of COVID-19. Journal of Education and e-learning Research, 7(3), 285-292.

Badmus, O. T., & Jita, L. (2023). Does further-mathematics performance predict student's achievement in Physics? A correlational diagnosis in senior secondary schools. JOTSE, 13(1), 381-392.

Bai, B., & Zang, X. (2025). Bilingual learning motivation and engagement among students in Chinese-English bilingual education programmes in Mainland China: competing or coexistent?. Journal of Multilingual and Multicultural Development, 1-19.

Bai, C. E., Chi, W., & Qian, X. (2014). Do college entrance examination scores predict undergraduate GPAs? A tale of two universities. China Economic Review, 30, 632-647.

Bala, B., & Behal, S. (2024, October). A Brief Survey of Data Preprocessing in Machine Learning and Deep Learning Techniques. In 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 1755-1762). IEEE.

Bandula, A. G. A., Wickramsinghe, Y. M., & Niranjala, S. A. U. (2024). Demographic and socio-economic determinants of alcohol consumption in sri lanka: a case study of anuradhapura district. 3rd International Research Symposium on Management IRSM, 237.

Bao, L., Wang, Y. T., Zhuang, J. L., Liu, A. J., Dong, Y. J., Chu, B., Chen, X. H., Lu, M. Q., Shi, L., Gao, S., Fang, L. J., Xiang, Q. Q., & Ding, Y. H. (2022). Machine Learning–

Based Overall Survial Prediction of Elderly Patients With Multiple Myeloma From Multicentre Real-Life Data. Frontiers in Oncology, 12, 922039.

Bao, W., & Jin, H., H. (2020). The Impact of New College Entrance Examination Reform on Academic Adaptation of Freshmen: Suppression or Enhancement?. Best Evidence in Chinese Education, 6(2), 883-885.

Barton, B., & Neville-Barton, P. (2003, February). Investigating the relationship between English language and mathematical learning. In Proceedings of the Third Conference of the European Society for Research in Mathematics Education, 28 February-3 March (pp. 1-10).

Basnyat, S. (2025). Parental Involvement on Children's Success: An Exploration of School Administration's Role from parental perspectives. Innovative Research Journal, 4(1), 211-229.

Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. Education and Information Technologies, 28(1), 905-971.

Behnke, M., Briner, N., Cullen, D., Schwerdtfeger, K., Warren, J., Basnet, R., & Doleck, T. (2021). Feature engineering and machine learning model comparison for malicious activity detection in the dns-over-https protocol. IEEE Access, 9, 129902-129916.

Ben Said, M., Hadj Kacem, Y., Algarni, A., & Masmoudi, A. (2023). Early prediction of Student academic performance based on Machine Learning algorithms: A case study of bachelor's degree students in KSA. Education and Information Technologies, 1-24.

Berberoğlu, G., & Tansel, A. (2014). Does private tutoring increase students' academic performance? Evidence from Turkey. International Review of Education, 60(5), 683-701.

Berhanu, F., & Abera, A. (2015). Students' performance prediction based on their academic record. International Journal of Computer Applications, 131(5), 27-35.

Berkman, B. E., Miller, W. K., & Grady, C. (2018). Is it ethical to use genealogy data to solve crimes?. Annals of internal medicine, 169(5), 333-334.

Bertolini, R., Finch, S. J., & Nehm, R. H. (2021). Enhancing data pipelines for predicting student performance: integrating feature selection with cross-validation. International Journal of Educational Technology in Higher Education, 18, 1-23.

Bhegade, K. B., & Shinde, S. V. (2016). Student performance prediction system with educational data mining. International Journal of Computer Applications, 146(5), 32-35.

Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with Decision Tree using CODAP and Jupyter Notebooks. Teaching Statistics, 43, S133-S142.

Biglan, A. (1973). The characteristics of subject matter in different academic areas. Journal of Applied Psychology, 57, 195-203.

Bilal, M., Omar, M., Anwar, W., Bokhari, R. H., & Choi, G. S. (2022). The role of demographic and academic features in a student performance prediction. Scientific Reports, 12(1), 12508.

Binder, T., Sandmann, A., Sures, B., Friege, G., Theyssen, H., & Schmiemann, P. (2019). Assessing prior knowledge types as predictors of academic achievement in the introductory phase of biology and physics study programmes using logistic regression. International Journal of STEM Education, 6, 1-14.

Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision Tree: from efficient prediction to responsible AI. Frontiers in Artificial Intelligence, 6.

Bolt, T., Nomi, J. S., Bzdok, D., & Uddin, L. Q. (2021). Educating the future generation of researchers: A cross-disciplinary survey of trends in analysis methods. PLoS biology, 19(7), e3001313.

Booker, L., & Holter, J. (2025). Research in Partnership: A Collection of Studies From the Tennessee Education Research Alliance. Peabody Journal of Education, 100(2), 133-138.

Borah, M. (2021). Motivation in learning. Journal of Critical Reviews, 8(2), 550-552.

Borin, N., Lindsey-Mullikin, J., & Krishnan, R. (2023). Understanding Consumers' Motivations and Product Preferences for Deconsumption. CSU Journal of Sustainability and Climate Change, 2(1), 4.

Bottomley, E., Kohnle, A., Mavor, K., Miles, P. J., & Wild, V. (2022). The relationship between gender and academic performance in undergraduate physics students: the role of physics identity, perceived recognition, and self-efficacy. European Journal of Physics, 44(2), 025701.

Brecht, A. A., & Burnett, D. D. (2019). Advising student-athletes for success: Predicting the academic success and persistence of collegiate student-athletes. NACADA Journal, 39(1), 49-59.

Bronfenbrenner, U. (1981). The ecology of human development: Experiments by nature and design (1st ed.). Cambridge: Harvard University Press.

Brown, G. T. (2024, September). Insights in assessment, testing, and applied measurement: 2022. In Frontiers in Education (Vol. 9, p. 1488012). Frontiers Media SA.

Brown, G. T. (2024, September). Insights in assessment, testing, and applied measurement: 2022. In Frontiers in Education (Vol. 9, p. 1488012). Frontiers Media SA.

Cadersaib, Z., Sta, H. B., & Gobin-Rahimbux, B. (2022, March). Enterprise Resource Planning integrated with Business Analytics in Higher Education. In 2022 IEEE Global Engineering Education Conference (EDUCON) (pp. 2114-2120). IEEE.

Cagliero, L., Canale, L., Farinetti, L., Baralis, E., & Venuto, E. (2021). Predicting student academic performance by means of associative classification. Applied Sciences, 11(4), 1420.

Cai, H., Jiang, Z., Chen, X., & Bailey, J. L. (2019). How to successfully train a modern nephrologist: Experience from US fellowship training practice. Kidney Diseases, 5(4), 204-210.

Calderon-Valenzuela, J., Payihuanca-Mamani, K., & Bedregal-Alpaca, N. (2022). Educational Data Mining to Identify the Patterns of Use made by the University Professors of the Moodle Platform. International Journal of Advanced Computer Science and Applications, 13(1), 321-328.

Cameron, C. E., Grimm, K. J., Steele, J. S., Castro-Schilo, L., & Grissmer, D. W. (2015). Nonlinear Gompertz curve models of achievement gaps in mathematics and reading. Journal of Educational Psychology, 107(3), 789.

Cano, F., García, Á., Berbén, A. B. G., & Justicia, F. (2014). Science learning: A path analysis of its links with reading comprehension, question-asking in class and science achievement. International Journal of Science Education, 36(10), 1710-1732.

Cao, Z. (2022). Research on the prediction and evaluation method of student performance in colleges and universities based on deep learning. Information Record Material (11), 123-125. (in Chinese) doi:10.16009/j.cnki.cn13-1295/tq.2022.11.050.

Caspari-Sadeghi, S. (2022, January). Applying learning analytics in online environments: Measuring learners' engagement unobtrusively. In Frontiers in Education (Vol. 7, p. 840947). Frontiers.

Caviola, S., Toffalini, E., Giofrè, D., Ruiz, J. M., Szűcs, D., & Mammarella, I. C. (2022). Math performance and academic anxiety forms, from sociodemographic to cognitive aspects: A meta-analysis on 906,311 participants. Educational Psychology Review, 1-37.

Chaka, C. (2021). Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: An overview of reviews.

Chakrabarty, P., Halder, K., & Rao, P. (2023). Tools and Methods of Educational Data Mining: A Review. Easy Chair Preprint, No. 9763.

Chan, C. K. Y., & Zhou, W. (2023). Deconstructing Student Perceptions of Generative AI (GenAI) through an Expectancy Value Theory (EVT)-based Instrument. arXiv preprint arXiv:2305.01186.

Chan, C., & Bray, M. (2014). Marketized private tutoring as a supplement to regular schooling: Liberal Studies and the shadow sector in Hong Kong secondary education. Journal of Curriculum Studies, 46(3), 361-388.

Chango, W., Lara, J. A., Cerezo, R., & Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12(4), e1458.

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01), 20-28.

Chatzimparmpas, A., Martins, R. M., & Kerren, A. (2023). VisRuler: Visual analytics for extracting decision rules from bagged and boosted Decision Tree. Information Visualization, 22(2), 115-139.

Chen, F. (2022). Prediction, Monitoring, and Management of the Classified Training Quality of English Majors Based on Support Vector Machine. International Journal of Emerging Technologies in Learning (Online), 17(24), 233.

Chen, J., Li, D., Huang, R., Chen, Z., & Li, W. (2023). Aero-engine remaining useful life prediction method with self-adaptive multimodal data fusion and cluster-ensemble transfer regression. Reliability Engineering & System Safety, 234, 109151.

Chen, Q. Z., Schnabel, T., Nushi, B., & Amershi, S. (2022). HINT: Integration Testing for AI-based features with Humans in the Loop. In 27th International Conference on Intelligent User Interfaces (pp. 549-565).

Chen, Q., Pei, C., Guo, Y., & Zhai, S. (2023). What drives academic peer effects in middle school classrooms in China: Peer composition or peer performance?. Heliyon, 9(6).

Chen, S., & Ding, Y. (2023). A Machine Learning Approach to Predicting Academic Performance in Pennsylvania's Schools. Social Sciences, 12(3), 118.

Chen, S., & Feng, Y. (2020). Literature Review on English Reform in China's new college entrance examination. Frontiers in Educational Research, 3(4), 47-51.

Chen, W., Sharifrazi, D., Liang, G., Band, S., Chau, K., & Mosavi, A. (2022). Accurate discharge coefficient prediction of streamlined weirs by coupling linear regression and deep convolutional gated recurrent unit. Engineering Applications of Computational Fluid Mechanics, 16(1), 1-19.

Chen, X. F., & Xue, P. (2020, April). A Study on The Admission Policy of College Entrance Examination for Ethnic Minorities in China from the Perspective of Cultural Analysis. In Proceedings of the 2020 2nd International Conference on Big Data and Artificial Intelligence (pp. 210-214).

Chen, X., & Li, Y. (2008). Language proficiency and mathematics learning. School Science and Mathematics, 108(3), 90-94.

Chen, Y. L. & Zhu. G. T. (2019). Influence of the Reform of College Entrance Examination on High School Physics Curriculum in Shanghai. Research in Education Assessment and Learning. 4(1).

Chen, Y., Jiang, M., & Kesten, O. (2020). An empirical evaluation of Chinese college admissions reforms through a natural experiment. Proceedings of the National Academy of Sciences, 117(50), 31696-31705.

Chen, Y., Li, Y., & Tian, J. (2022, November). New College Entrance Examination Reform in the Zhejiang Province of China. In 2022 International Conference on Science Education and Art Appreciation (SEAA 2022) (pp. 1028-1036). Atlantis Press.

Chen, Z., Li, W., & Zhan, G. (2023). A study of social mobility under the Imperial Examination System in Qing Dynasty based on Bourdieur's theories. In SHS Web of Conferences (Vol. 180, p. 01003). EDP Sciences.

Cheng, L., Li, M., Kirby, J. R., Qiang, H., & Wade-Woolley, L. (2010). English language immersion and students' academic performance in English, Chinese and mathematics. Evaluation & Research in Education, 23(3), 151-169.

Cheng, Y. Y. (2020). Academic self-efficacy and assessment. Educational Psychology, 40(4), 389-391.

Chong, S. W., Isaacs, T., & McKinley, J. (2023). Ecological systems theory and second language research. Language Teaching, 56(3), 333-348.

Christensen, D., Taylor, C. L., Hancock, K. J., & Zubrick, S. R. (2022). School readiness is more than the child: a latent class analysis of child, family, school and community aspects of school readiness. Australian Journal of Social Issues, 57(1), 125-143.

Chui, K. T., Liu, R. W., Zhao, M., & De Pablos, P. O. (2020). Predicting students' performance with school and family tutoring using generative adversarial network-based deep support vector machine. IEEE Access, 8, 86745-86752.

Cimini, C., Boffelli, A., Lagorio, A., Kalchschmidt, M., & Pinto, R. (2020). How do industry 4.0 technologies influence organisational change? An empirical analysis of Italian SMEs. Journal of Manufacturing Technology Management, 32(3), 695-721.

Cimmiyotti, C. B. (2013). Impact of Reading Ability on Academic Performance at the Primary Level. Graduate Master's Theses, Capstones, and Culminating Projects. 127.

Cleopas, B. C., & Onwuchekwa, F. C. (2024). Impact of chemistry teachers' qualification and years of experience on academic performance of secondary school chemistry students in Ogbia local government area Bayelsa state. Contemporary Research in Education and English Language Teaching, 6(1), 62-72.

Coelho, O. B., & Silveira, I. (2017). Deep learning applied to learning analytics and educational data mining: A systematic literature review. In Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)，28, (1), 143.

Coldwell, J., Craig, A., & Goold, A. (2011). Using etechnologies for active learning. Interdisciplinary Journal of Information, Knowledge, and Management, 6(3), 95-106.

Cortés Pascual, A., Moyano Muñoz, N., & Quilez Robres, A. (2019). The relationship between executive functions and academic performance in primary education: Review and meta-analysis. Frontiers in psychology, 10, 1582.

Costabile, M., Simpson, B., Turkanovic, J., & Hughes, B. P. (2024). Enhancing teaching effectiveness in biochemistry labs: Author reflections and improvement strategies. Biochemistry and Molecular Biology Education, 52(5), 559-568.

Creech, L. R., & Sweeder, R. D. (2012). Analysis of student performance in large-enrolment life science courses. CBE—Life Sciences Education, 11(4), 386-391.

Crosnoe, R. (2009). Family–school connections and the transitions of low-income youths and english language learners from middle school to high school. Developmental Psychology, 45(4), 1061.

Cuéllar-Rojas, O. A., Hincapié-Montoya, M., Contero, M., & Güemes-Castorena, D. (2022, December). Bibliometric analysis and systematic literature review of the intelligent tutoring systems. In Frontiers in Education (Vol. 7, p. 1047853). Frontiers.

Cui, J., Lv, L., Du, H., Cui, Z., & Zhou, X. (2022). Language ability accounts for ethnic difference in mathematics achievement. Frontiers in Psychology, 13, 929719.

Dahdouh K, Dakkak A, Oughdir L, et al. Improving online education using big data technologies[J]. The role of technology in education, 2020: 15.

Daumiller, M., Gaspard, H., Dickhäuser, O., & Dresel, M. (2025). Bridging teacher motivation and instruction: Relevance of student - oriented goals for teaching alongside personal achievement goals and self‐efficacy. British Journal of Educational Psychology.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. Journal of family psychology, 19(2), 294.

De Menezes, M. F., Netto, J. F. D. M., & Lopes, A. M. M. (2022, October). Educational Data Mining: Analysis Based On an Intelligent Tutoring System for Teaching Algebra. In 2022 IEEE Frontiers in Education Conference (FIE) (pp. 1-7). IEEE.

Deepa, S. M., Revathi, N., Sivakami, K., & Pareek, P. K. (2022, October). Machine Learning based Education System with Sentiment Analysis for Students. In 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon) (pp. 1-6). IEEE.

Dejing, K., Jianlin, Y., Xiaoying, Y., & Xianming, Q. (2017). A Study on the Issues of Conducting China’s Manufacturing Power Strategy. Strategic Study of Chinese Academy of Engineering, 19(3), 6-12.

Deng, Q. (2024). The Impact of Interpersonal Relationships on Students' Academic Achievement in Primary and Secondary Schools. Journal of Education, Humanities and Social Sciences, 29, 49-55.

Depren, S. K., Aşkın, Ö. E., & Öz, E. (2017). Identifying the classification performances of educational data mining methods: A case study for TIMSS. Educational Sciences: Theory & Practice, 17(5).

Dermentzi, E., Zotou, M., Tambouris, E., & Tarabanis, K. (2022). Using the problem based learning method and educational technologies to teach open data: A design-based research approach. Education and Information Technologies, 27(6), 8859-8882.

Ding, G. Y., Qin, X. G., & Wang, X. (2019). Research on prediction of college students' performance in laboratory courses based on educational data mining. China Education Informatisation(13),41-44. (in Chinese)

Disha, R. A., & Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity, 5(1), 1.

Dol, S. M., & Jawandhiya, P. M. (2022, July). Use of Data mining Tools in Educational Data Mining. In 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT) (pp. 380-387). IEEE.

Dol, S. M., & Jawandhiya, P. M. (2023). Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining—A survey. Engineering Applications of Artificial Intelligence, 122, 106071.

Dol, S. M., & Jawandhiya, P. M. (2024, April). Application of Preprocessing Techniques on the Dataset obtained from Education Sector to Predict Students' Performance. In 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon) (pp. 1-6). IEEE.

Donis, K., Aikens, M. L., Swamy, U., Delgado, M., Gillespie, M., Graves, P., & Eddy, S. L. (2024). Learning Assistants and Instructors Provide Social Support That Influences Student Engagement Differently in Undergraduate Active Learning Chemistry Courses. Journal of Chemical Education, 101(8), 2989-3002.

Dorta‐Guerra, R., Marrero, I., Abdul‐Jalbar, B., Trujillo‐González, R., & Torres, N. V. (2019). A new academic performance indicator for the first term of first‐year science degrees students at la laguna university: A predictive model. FEBS open bio, 9(9), 1493-1502.

Doz, D., & Sliško, J. (2024). The cognitive reflection test and students' achievements in mathematics and physics. European Journal of Science and Mathematics Education, 12(1), 85-96.

Du, C., Qin, K., Wang, Y., & Xin, T. (2021). Mathematics interest, anxiety, self-efficacy and achievement: Examining reciprocal relations. Learning and Individual Differences, 91, 102060.

Du, X., Chen, Y., Zhang, X., & Guo, Y. (2022). Study on Feature Engineering and Ensemble Learning for Student Academic Performance Prediction. International Journal of Advanced Computer Science and Applications, 13(5), 495-502.

Duan, L., He, J., Li, M., Dai, J., Zhou, Y., Lai, F., & Zhu, G. (2021). Based on a decision tree model for exploring the risk factors of smartphone addiction among children and adolescents in China during the COVID-19 pandemic. Frontiers in Psychiatry, 12, 652356.

Duan, W., Guan, Y., & Bu, H. (2018). The effect of parental involvement and socioeconomic status on junior school students' academic performance and school behaviour in China. Frontiers in psychology, 9, 952.

Dunlap, M. R. (1998). Adjustment and developmental outcomes of students engaged in service learning. Journal of Experiential Education, 21(3), 147-153.

Dunn, T. J., & Kennedy, M. (2019). Technology Enhanced Learning in higher education; motivations, engagement and academic performance. Computers & Education, 137, 104-113.

Dupuis, J., Savick, S., & Fenster, M. (2020). Relationship between Self-Efficacy Measured by the TSES Scale and Teacher Participation in PDS Activity. Professional Educator, 43(1), 47-58.

Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., Cai, K., Clifford, R. M., Ebanks, C., Griffin, J. A., Henry G. T., Howes, C., Iriondo-Perez, J., Jeon, H., Mashburn, A. J., Peisner-Feinberg, E., Pianta, R. C., Vandergrift, N., & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. Child development, 78(2), 558-580.

Eccles-Parsons, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviours. Performance and performance motivation, 1983, 75-146.

Echazarra, A., & Radinger, T. (2019). Learning in rural schools: INSIGHTS FROM PISA, TALIS AND THE LITERATURE. OECD Education Working Papers, (196), 0_1-77.

Egara, F. O., & Mosimege, M. (2024). Effect of flipped classroom learning approach on mathematics achievement and interest among secondary school students. Education and Information Technologies, 29(7), 8131-8150.

Engels, N., de Graav, G. N., van der Nat, P., van den Dorpel, M., Stiggelbout, A. M., & Bos, W. J. (2022). Shared decision-making in advanced kidney disease: a scoping review. BMJ open, 12(9), e055248.

Engida, M. A., Iyasu, A. S., & Fentie, Y. M. (2024, July). Impact of teaching quality on student achievement: student evidence. In Frontiers in Education (Vol. 9, p. 1367317). Frontiers Media SA.

Ewere, C. R. (2023). Effects of self-efficacy on the academic achievement of biology students in secondary schools in Delta State. Journal of Advances in Education and Philosophy, 7(4), 110-114.

Fadlilah, A. N. (2022). Technology Challenge: EFL Teacher Experience Teaching Online at Kindergarten. Jurnal Pendidikan Anak Usia Dini Undiksha, 10(2), 228-238.

Fakapulia, I. F., Samalia, L., & Wibowo, E. (2023). Social factors are associated with academic stress in Pasifika students at the University of Otago. Waikato Journal of Education, 28(1).

Fan, J., Wade, J. W., Bian, D., Key, A. P., Warren, Z. E., Mion, L. C., & Sarkar, N. (2015). A Step towards EEG-based brain computer interface for autism intervention. In 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 3767-3770). IEEE.

Fan, Y., & Cheng, C. (2017). Examining trends in Chinese secondary education: From test-oriented to comprehensive academic and personal student development. US-China Education Review, 7(10), 453-462.

Farley, A., & Yang, H. H. (2020). Comparison of Chinese Gaokao and Western university undergraduate admission criteria: Australian ATAR as an example. Higher Education Research & Development, 39(3), 470-484.

Fearnley, M., Malay, C., & Fabia, J. N. (2022). Factors Affecting Student Satisfaction Perceived Learning and Academic Performance in an Emergency Online Science Course. International Journal of Information and Education Technology, 12(12), 1354-1362.

Feng, E., & Li, C. (2024). Examining the role of achievement goals in L2 learning: A cross-sectional and longitudinal study. System, 125, 103438.

Feng, S., Wong, Y. K., Wong, L. Y., & Hossain, L. (2019). The Internet and Facebook usage on academic distraction of college students. Computers & Education, 134, 41-49.

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. Journal of business research, 94, 335-343.

Fernández, M., Rezaei, S., Rezaei Mianroodi, J., Fritzen, F., & Reese, S. (2020). Application of artificial neural networks for the prediction of interface mechanics: a study on grain boundary constitutive behaviour. Advanced modelling and simulation in engineering sciences, 7(1), 1-27.

Flogie, A., & Krabonja, M. V. (2023, June). Artificial intelligence in education: developing competencies and supporting teachers in implementing AI in school learning environments. In 2023 12th Mediterranean Conference on Embedded Computing (MECO) (pp. 1-6). IEEE.

Flores, V., Heras, S., & Julian, V. (2022). Comparison of predictive models with balanced classes using the SMOTE method for the predict of student dropout in higher education. Electronics, 11(3), 457.

Frame, M. L. (2020). The Institutional Pillars of China's National College Entrance Exam: A Case Study of Gaozhong High School and NCEE Reforms.

Fu, R., Lee, J., Chen, X., & Wang, L. (2020). Academic self‐perceptions and academic achievement in Chinese children: A multiwave longitudinal study. Child Development, 91(5), 1718-1732.

Fu, X., Hong, S., & Wang, T. (2023, June). Evaluation Model Optimization and Algorithm Design for Financial Transformation of Medicine Business Enterprises based on Data Mining. In Proceedings of the 3rd International Conference on Big Data Economy and Information Management, BDEIM 2022, December 2-3, 2022, Zhengzhou, China.

Fu, Y. (2024). The Impact of Gaokao High-Stakes Testing on Student Mental Health in China: An Analysis of Stress Levels and Coping Mechanisms Among Senior High School Students. Research and Advances in Education, 3(5), 23-32.

Gafoor, A. K., & Ashraf, M. (2016). Academic Self-Efficacy Scale-2006. Calicut: Department of Education, University of Calicut. DOI, 10.

Galikyan, I., & Admiraal, W. (2019). Students' engagement in asynchronous online discussion: The relationship between cognitive presence, learner prominence, and academic performance. The Internet and Higher Education, 43, 100692.

Galvan, J. A., Casman, E., Fisher, E., Nair, I., & Small, M. J. (2014). Assessing the role of 21st century skills on internship outcomes in a steel multinational enterprise. In 2014 ASEE Annual Conference & Exposition (pp. 24-209).

Gao, H., Hashim, H., & Md Yunus, M. (2024). The effect of project-based learning (PjBL) Class Model on Chinese undergraduates' argumentative writing performance. Cogent Education, 11(1), 2396715.

Gao, J., Yang, Z., Li, F., Yasen, B., & Wen, S. (2023). The influence of cognitive ability in Chinese reading comprehension: can working memory updating change Chinese primary school students' reading comprehension performance?. Frontiers in Psychology, 14, 1283781.

Gao, N. (2023). The Influence of College Entrance Examination Policy on Educational Equity in China. BCP Business & Management, 41, 59–63.

Gao, P., Wu, W., & Li, J. (2021). Multi-source fast transfer learning algorithm based on support vector machine. Applied Intelligence, 51, 8451-8465.

Garg, A., Lilhore, U. K., Ghosh, P., Prasad, D., & Simaiya, S. (2021, August). Machine Learning-based Model for Prediction of Student's Performance in Higher Education. In 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 162-168). IEEE.

George, A. H., Shahul, A., George, A. S., Baskar, T., & Hameed, A. S. (2023). A Survey Study on Big Data Analytics to Predict Diabetes Diseases Using Supervised Classification Methods. Partners Universal International Innovation Journal, 1(1), 1-8.

Getie, A. S. (2020). Factors affecting the attitudes of students towards learning English as a foreign language. Cogent Education, 7(1), 1738184.

Giraldo, M., & Rodríguez, J. (2019). Comparison of Statistical and Machine Learning models for pipe failure modelling in Water Distribution Networks (WDN). In Proceedings of the 4th International Electronic Conference on Water Sciences, 13–29 November 2019.

Gong, Y., X. (2020, December). Research on the Evolution of the College Entrance Examination Policy in Tianjin Based on the Theory of Multiple Streams. In 2020 3rd International Conference on Humanities Education and Social Sciences (ICHESS 2020) (pp. 118-125). Atlantis Press.

Gong, Y., Yuksel, T., Magana, A. J., & Bryan, L. A. (2015, June). Engineering and Physics Students' Perceptions about Learning Quantum Mechanics via Computer Simulations. In 2015 ASEE Annual Conference & Exposition (pp. 26-614).

Goodday, S. M., Rivera, D., Foran, H., King, N., Milanovic, M., Keown-Stoneman, C. D., Horrocks, J., Tetzlaff, E., Bowie, C. R., Pickett, W., Harkness, K., Saunders, K. E., Cunningham, S., McNevin, S., & Duffy, A. (2019). U-Flourish university students well-being and academic success longitudinal study: a study protocol. BMJ open, 9(8), e029854.

Gopal, R., Singh, V., & Aggarwal, A. (2021). Impact of online classes on the satisfaction and performance of students during the pandemic period of COVID 19. Education and Information Technologies, 26(6), 6923-6947.

Gordon, A. S., Plumblee, J. M., & Dancz, C. L. (2017). Developing Leadership through an Immersive Service-Oriented International Internship. Paper presented at 2017 ASEE Annual Conference & Exposition, Columbus, Ohio. 10.18260/1-2—28145.

Gore, J. S., Thomas, J., Jones, S., Mahoney, L., Dukes, K., & Treadway, J. (2016). Social factors that predict fear of academic success. Educational Review, 68(2), 155-170.

Gremmen, M. C., Dijkstra, J. K., Steglich, C., & Veenstra, R. (2017). First selection, then influence: Developmental differences in friendship dynamics regarding academic performance. Developmental psychology, 53(7), 1356.

Gresse von Wangenheim, C., Hauck, J. C., Pacheco, F. S., & Bertonceli Bueno, M. F. (2021). Visual tools for teaching machine learning in K-12: A ten-year systematic mapping. Education and Information Technologies, 26(5), 5733-5778.

Gu, Y. (2023). The Influence of Family Income Level on the Future Major Selection Approaches of Chinese Ordinary High School Students and the Potential Solutions. Studies in Social Science & Humanities, 2(5), 41-50.

Gumasing, M. J. J., & Castro, F. M. F. (2023). Determining ergonomic appraisal factors affecting the learning motivation and academic performance of students during online classes. Sustainability, 15(3), 1970.

Gumba, G., & Paragas, J. R. (2022, September). Prediction Analysis Of Student Admission To Information Technology Education (ITE) Programs Using Classification Algorithm. In 2022 Fifth International Conference on Vocational Education and Electrical Engineering (ICVEE) (pp. 112-117). IEEE.

Guo, C., Guo, M., & Hao, X. (2021). Do Students Prioritize Majors or Specific Colleges? Analyzing the Factors That Influence Preferences in China. Sustainability, 13(16), 9196.

Guo, F., Duan, Y., He, S., Gong, Q., & Yao, Z. (2022). Analysis of sustainable development questions of college entrance examination geography papers in China: 2010–2020. Sustainability, 14(3), 1526.

Guo, M., & Yang, X. (2021, October). Analysis and Computer Intelligent Prediction Model of College Entrance Examination Number through ARIMA and Big Data Technology. In 2021 IEEE International Conference on Data Science and Computer Application (ICDSCA) (pp. 144-147). IEEE.

Guo, Z., Qi, C., Yang, J., Xu, Y., & Li, S. (2023). How family structure influences middle-school students' involvement in physical exercise and their academic performance in China. Humanities and Social Sciences Communications, 10(1), 1-9.

Guy-Evans, O. (2024). Bronfenbrenner's ecological systems theory. Simply Psychology. Retrieved from https://www.simplypsychology.org/bronfenbrenner.html.

Hagemann, L., & Abramova, O. (2023). Sentiment, we-talk and engagement on social media: Insights from Twitter data mining on the US presidential elections 2020. Internet Research.

Hajovsky, D. B., Oyen, K. A., Chesnut, S. R., & Curtin, S. J. (2020). Teacher–student relationship quality and math achievement: The mediating role of teacher self‐efficacy. Psychology in the Schools, 57(1), 111-134.

Hamid, S., & Singaram, V. S. (2016). Motivated strategies for learning and their association with academic performance of a diverse group of 1st-year medical students. African Journal of Health Professions Education, 8(1), 104-107.

Hamoud, A., & Humadi, A. (2019). Student’s success prediction model based on artificial neural networks (ANN) and a combination of feature selection methods. Journal of Southwest Jiaotong University, 54(3).

Hamoud, A., Hashim, A. S., & Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. International Journal of Interactive Multimedia and Artificial Intelligence, 5, 26-31.

Han, N. (2022). Educational Change Case Study: from the Educational Leadership Perspective to Solve the Problem of Shandong Province’s College Entrance Examination Policy. International Journal of New Developments in Education, 4(10), 59-62.

Han, S., & Fu, H. (2022). Experimentation, policy learning, and China’s education reforms. International Journal of Educational Research, 116, 102057.

Han, Z. (2024). Academic stress and coping strategies of Chinese high school students-a qualitative investigation. Advances in Education, Humanities and Social Science Research, 12(1), 270-270.

Harveson, A. T., Hannon, J. C., Brusseau, T. A., Podlog, L., Papadopoulos, C., Hall, M. S., & Celeste, E. (2019). Acute Exercise and Academic Performance in Middle School Students. IJERPH, 16(19), 1-7.

Hasib, K. M., Rahman, F., Hasnat, R., & Alam, M. G. R. (2022, January). A machine learning and explainable ai approach for predicting secondary school student performance. In 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0399-0405). IEEE.

Hassan, L. F. A., Demong, N. A. R., Salleh, M. Z. M., Omar, E. N., & Alwi, A. (2022). Feature Selection of Digital Addiction Prediction Among Students Using WEKA. Global Business and Management Research, 14(1), 70-81.

Hayat, A. A., Shateri, K., Amini, M., & Shokrpour, N. (2020). Relationships between academic self-efficacy, learning-related emotions, and metacognitive learning strategies with academic performance in medical students: a structural equation model. BMC medical education, 20(1), 1-11.

He, B., & Thompson, C. (2022). Family involvement and English learners' outcomes: A synthetic analysis. International Review of Education, 68(3), 409-440.

Held, K. A. S. (2023). School-Based Agricultural Education Student Perceptions of their Motivation through the lens of Situated Expectancy-Value Theory (Doctoral dissertation, The Ohio State University).

Henry, D. L., Baltes, B., & Nistor, N. (2014). Examining the relationship between math scores and English language proficiency. Journal of Educational Research and Practice, 4(1), 2.

Hidayatullah, A., & Csíkos, C. (2024). The role of students' beliefs, parents' educational level, and the mediating role of attitude and motivation in students' mathematics achievement. The Asia-Pacific Education Researcher, 33(2), 253-262.

Hilmi, M. F., Mahmud, N., Harizan, S. H. M., & Mustapha, Y. (2022). DEEP LEARNING TECHNIQUE FOR DETECTING ONLINE LEARNING BEHAVIOUR. International Journal of Education and Pedagogy, 4(3), 12-19.

Ho, C. M., Yeh, C. C., Wang, J. Y., Hu, R. H., & Lee, P. H. (2021). Curiosity in online video concept learning and short-term outcomes in blended medical education. Frontiers in medicine, 8, 772956.

Holyoke, L., & Fein, A. H. (2009). Making Sense of Schools: Perceptions of Student Teacher Interns. Northwest Journal of Teacher Education, 7(1), 7.

Holzinger, A. (2019). Introduction to machine learning & knowledge extraction (MAKE). Machine Learning and Knowledge Extraction, 1(1), 1-20.

Hong, C. W., Lee, C., Lee, K., Ko, M. S., Kim, D. E., & Hur, K. (2020). Remaining useful life prognosis for turbofan engine using explainable deep neural networks with dimensionality reduction. Sensors, 20(22), 6626.

Howlett, Z. M. (2022). Performative secularism: school-sponsored prayer in China's National College Entrance Exam. Critical Asian Studies, 54(3), 441-469.

Hu, C., Ma, Y., & Chen, T. (2021). Application on Online Process Learning Evaluation Based on Optimal Discrete Hopfield Neural Network and Entropy Weight TOPSIS Method. Complexity, 2021, 1-9.

Hu, J. X. (2018). On the reform of high school level examination under the background of the new college entrance examination reform (Part 2). Hubei Entrance Examination, 1, 10–15.

Hu, L. (2022). Opportunities and Challenges: An Examination of Educational Landscape for People with Visual Impairment in China. Current Issues in Comparative Education, 24(1), 96-113.

Huang, F., Hoi, C. K. W., & Teo, T. (2018). The influence of learning style on English learning performance among undergraduates in Mainland China. Journal of Psycholinguistic Research, 47(5), 1069-1084.

Huang, J. (2020). Operation mechanism and evaluation of "county high school education model" in the context of Chinese college entrance examination system. Science Insights Education Frontiers, 7(2), 879-891.

Huang, J., Hu, X., & Yang, F. (2011). Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker. Measurement, 44(6), 1018-1027.

Huang, L.M., & Zhong, B.L. (2021). Institutional analysis of "cold weather" in the new college entrance examination-taking physics as an example. Journal of Education, 17(01), 100–109.

Huangfu, Q., Wei, N., Zhang, R., Tang, Y., & Luo, G. (2023). Social support and continuing motivation in chemistry: the mediating roles of interest in chemistry and chemistry self-efficacy. Chemistry Education Research and Practice, 24(2), 478-493.

Hung, H. C., Liu, I. F., Liang, C. T., & Su, Y. S. (2020). Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education. Symmetry, 12(2), 213.

Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. Annals of data science, 10(3), 637-655.

Hussain, S., Ayoub, M., Jilani, G., Yu, Y., Khan, A., Wahid, J. A., Butt, M. F. A., Yang, G. Q., Moller, D. P. F., & Weiyan, H. (2022). Aspect2Labels: A novelistic decision support system for higher educational institutions by using multi-layer topic modelling approach. Expert Systems with Applications, 209, 118119.

Hussain, S., Dahan, N. A., Fadl, B. M., & Ribata, N. (2018). Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. International Journal of Electrical and Computer Engineering (IJECE), 9(2), 447-459.

Huynh-Cam, T. T., Chen, L. S., & Huynh, K. V. (2022). Learning performance of international students and students with disabilities: early prediction and feature selection through educational data mining. Big Data and Cognitive Computing, 6(3), 94.

Hwang, Y. (2023). When makers meet the metaverse: Effects of creating NFT metaverse exhibition in maker education. Computers & Education, 194, 104693.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. Psychological bulletin, 107(2), 139.

Ihrmark, D., & Tyrkkö, J. (2023). Learning text analytics without coding? An introduction to KNIME. Education for Information, (Preprint), 1-17.

Imam, O. A., Abas-Mastura, M., & Jamil, H. (2013). Correlation between reading comprehension skills and students' performance in mathematics. International Journal of Evaluation and Research in Education (IJERE), 2(1), 1-8.

Intakaew, K., & Wattanakasiwich, P. (2022, November). Analysis of kinematics graph interpretation skills using RapidMiner. In Proceedings of the IUPAP International Conference on Physics Education 2022 (pp. 102-102).

Inusah, F., Missah, Y. M., Najim, U., Twum, F. (2022). Data Mining and Visualisation of Basic Educational Resources for Quality Education. International Journal of Engineering Trends and Technology, 70 (12), 296-307.

Islam, A., & Tasnim, S. (2021). An Analysis of Factors Influencing Academic Performance of Undergraduate Students: A Case Study of Rabindra University, Bangladesh (RUB). Shanlax International Journal of Education, 9(3), 127-135.

Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. Decision Analytics Journal, 100204.

Jacobson, L. O. (2000). Editor's choice: Valuing diversity—Student teacher relations that enhance performance. Community College Review, 28(1), 49-66.

Jahantabi Neghad, S., Makvandi, F., Kiani, E. A., Amirnegad, G., & Chenari, V. (2025). Presenting the conceptual model of organizational democracy with an emphasis on the development of employees' psychological capital. Management and Educational Perspective, 6(4), 165-182.

Jamal, B., Niazi, S., Arshad, R., Muhammad, W., Manzoor, F., Khan, M. H. N., Riaz, A., & Naseem, A. (2023). Addressing gender disparities in education: Empowering girls through education in Pakistan. Russian Law Journal, 11(12s).

Jebu, S. R., & Raddam, Y. U. (2024). Impact of using strategy of Team Pair Solo to develop the trend towards History for female students in the fourth grade. Nasaq, 42(9).

Jenal, N., Taib, S. A., Iliyas, S. M. M., Sa'adan, N., Saleh, N. S., & Noorezam, M. (2022). Investigating Students' Learning Motivation Based on Value, Expectancy, and Affective Components. International Journal of Academic Research in Business and Social Sciences, 12(10), 641-661.

Jerrim, J., Prieto-Latorre, C., Marcenaro-Gutierrez, O. D., & Shure, N. (2025). Teacher Self-Efficacy, Instructional Practice, and Student Outcomes: Evidence from the TALIS Video Study. American Educational Research Journal, 62(2), 378-413.

Jiang, H. (2023). The Impact of Motivation on Academic Performance of Chinese College Students. In SHS Web of Conferences (Vol. 180, p. 04011). EDP Sciences.

Jiang, H. (2025). Deep Learning Based Personalized English Listening Learning Path Recommendation Algorithm. Systems and Soft Computing, 200210.

JIANG, Y., & LI, A. J. (2018). COPUS-based Observation and Analysis on Chinese and American College Classroom. 2018 International Conference on Education Reform, Management and Applied Social Science (ERMASS 2018).

Jing, X., & Liu, L. (2019). The Equity of Gaokao (National University/College Entrance Examination) in China. Review of Educational Theory, 2(3), 29-34.

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' performance. Teachers college record, 114(10), 1-39.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. Developmental psychology, 45(3), 850-867.

Jung, H. (2025). Schooling or parental involvement? Assessing the impact and mechanism of private elementary school on academic achievement. The Japanese Economic Review, 1-18.

Kadaruddin, K. (2023). Empowering Education through Generative AI: Innovative Instructional Strategies for Tomorrow's Learners. International Journal of Business, Law, and Education, 4(2), 618-625.

Kakish, H., & Al-Eisawi, D. (2023). A Predictive Model for Assessing Satisfaction with Online Learning for Higher Education Students During and After COVID-19 Using Data Mining and Machine Learning Techniques: A Case of Jordanian Institutions. In Proceedings of the 25th International Conference on Enterprise Information Systems (ICEIS 2023) - Volume 1, pages 156-163.

Kamogashira, T. (2024). Re-examining the Analysis of the Compiling Trends of Local Government History and Geography Published in Shizuoka Prefecture Using "Chosen Tsushin-shi (Korean Envoy to Japan)" as a Keyword., Educational Research (5), 309-318.

Karalić, I., Pantelić, O., & Đukić, M. (2023). Comparative analysis of open-source data mining tools. The Scientific Heritage, 106, 55–61.

Karbownik, M. S., Kręczyńska, J., Kwarta, P., Cybula, M., Wiktorowska-Owczarek, A., Kowalczyk, E., Pietras, T., & Szemraj, J. (2020). Effect of Supplementation with Saccharomyces Boulardii on Academic Examination Performance and Related Stress in Healthy Medical Students: A Randomized, Double-Blind, Placebo-Controlled Trial. Nutrients, 12(5), E1469-E1469.

Kaspi, S., & Venkatraman, S. (2023). Data-Driven Decision-Making (DDDM) for Higher Education Assessments: A Case Study. Systems, 11(6), 306.

Katrina, W., Damanik, H. J., Parhusip, F., Hartama, D., Windarto, A. P., & Wanto, A. (2019, August). C. 45 classification rules model for determining students level of understanding of the subject. In Journal of Physics: Conference Series (Vol. 1255, No. 1, p. 012005). IOP Publishing.

Kaur, T., McLoughlin, E., & Grimes, P. (2022). Mathematics and science across the transition from primary to secondary school: a systematic literature review. International Journal of STEM Education, 9(1), 13.

Kaveh, M., & Mesgari, M. S. (2022). Application of meta-heuristic algorithms for training neural networks and deep learning architectures: A comprehensive review. Neural Processing Letters, 1-104.

Ke, S. (2025). Chinese morphological awareness assessment and its relation to reading acquisition: a cross-cultural meta-analysis. Humanities and Social Sciences Communications, 12(1), 1-13.

Kekeba, S. K., Gure, A., & Olkaba, T. T. (2025). Effects of jigsaw learning strategy integrated with computer simulations on grade 10 students' achievement and attitude and their correlations in chemistry. Interactive Technology and Smart Education, 22(1), 134-162.

KG, S., Kurni, M., KG, S., & Kurni, M. (2021). Educational data mining & learning analytics. A Beginner's Guide to Learning Analytics, 29-60.

Khan, H., Gul, R., & Zeb, M. (2023). The Effect of Students' Cognitive and Emotional Engagement on Students' Academic Success and Academic Productivity. Journal of Social Sciences Review, 3(1), 322-334.

Khan, M. A., Abu-Khadrah, A., Siddiqui, S. Y., Ghazal, T. M., Faiz, T., Ahmad, M., & Lee, S. W. (2022). Support-Vector-Machine-based Adaptive Scheduling in Mode 4 Communication. Computers, Materials & Continua, 73(2).

Khan, R. U., Khan, A., Sadiq, N., & Uddin, S. (2025). A Review of Challenges Faced by Educators and their Ripple Effects on Learners' Outcomes. Dialogue Social Science Review (DSSR), 3(1), 231-245.

Kim, H. J., Hong, A. J., & Song, H. D. (2019). The roles of academic engagement and digital readiness in students' performances in university e-learning environments. International Journal of Educational Technology in Higher Education, 16(1), 1-18.

Kim, M., Shin, S. Y., Kang, M., Yi, B. K., & Chang, D. K. (2019). Developing a standardization algorithm for categorical laboratory tests for clinical big data research: retrospective study. JMIR medical informatics, 7(3), e14083.

Kirkham, J., Chapman, E., & Male, S. (2023). Measuring Motivation for Mathematics Course Choice in Secondary School Students: Interrelationships Between Cost and Other Situated Expectancy-Value Theory Components. SAGE Open, 13(2), 21582440231180671.

Kleszczewska, D., Dzielska, A., Salonna, F., & Mazur, J. (2018). The association between physical activity and general life satisfaction in lower secondary school students: The role of individual and family factors. Community mental health journal, 54, 1245-1252.

Kong, X. M. (2020). The characteristics and influence of China's new college entrance examination reform. In 2020 4th International Conference on Economics, Management Engineering and Education Technology. Suzhou, 340-342.

Kou, J. W., Fan, L. Y., Chen, H. C., Chen, S. Y., Hu, X., Zhang, K., Kovelman, I., & Chou, T. L. (2024). Neural substrates of l2-l1 transfer effects on phonological awareness in young chinese-english bilingual children. NeuroImage, 291, 120592.

Križanić, S. (2020). Educational data mining using cluster analysis and decision tree technique: A case study. SAGE Open, 10(1), 1-12.

Kumaran, M., Pham, T. M., Wang, K., Usman, H., Norris, C. M., MacDonald, J., Oudit G. Y., Saini, V., & Sikdar, K. C. (2022). Predicting the risk factors associated with severe outcomes among COVID-19 Patients–Decision tree modelling approach. Frontiers in Public Health, 10, 838514.

Kumari, A., Ortiz-Rodríguez, J. C., Mers-Noble, S. L., Enderle, B., & Velázquez, J. M. (2025). Enhancing Student Engagement and Success in General Chemistry Through Supplemental Co-Class Models. Journal of Chemical Education. https://doi.org/10.1021/acs.jchemed.3c01137

Kuo, Y. C., Chu, H. C., & Huang, C. H. (2015). A learning style-based grouping collaborative learning approach to improve EFL students' performance in English courses. Journal of Educational Technology & Society, 18(2), 284-298.

Kwedlo, W., & Czochanski, P. J. (2019). A hybrid MPI/OpenMP parallelization of $ K $-means algorithms accelerated using the triangle inequality. Ieee Access, 7, 42280-42297.

Lake, B. (2022). Technology and data analysis in the accounting profession. (Honors Thesis. Oklahoma State University).

Lara, L., & Saracostti, M. (2019). Effect of parental involvement on children's academic performance in Chile. Frontiers in psychology, 10, 1464.

Lau, K. L., & Qian, Q. (2024). Enhancing students' classical Chinese reading through a two-year flipped self-regulated learning intervention program. Reading and Writing, 1-35.

Lavelle-Hill, R., Frenzel, A. C., Goetz, T., Lichtenfeld, S., Marsh, H. W., Pekrun, R., Sakaki, M., Smith, G., & Murayama, K. (2024). How the predictors of math achievement change over time: A longitudinal machine learning approach. Journal of Educational Psychology, 116(8), 1383–1403.

Law, K. M., Geng, S., & Li, T. (2019). Student enrolment, motivation and learning performance in a blended learning environment: The mediating effects of social, teaching, and cognitive presence. Computers & Education, 136, 1-12.

Ledhem, M. A. (2022). Data mining techniques for predicting the financial performance of Islamic banking in Indonesia. Journal of Modelling in Management, 17(3), 896-915.

Lee, I., & Perret, B. (2022, June). Preparing High School Teachers to Integrate AI Methods into STEM Classrooms. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 11, pp. 12783-12791).

Lee, J., Song, H. D., & Hong, A. J. (2019). Exploring factors, and indicators for measuring students' sustainable engagement in e-learning. Sustainability, 11(4), 985.

Lee, S. H., Chang, H. J. J., & Zhao, L. (2023). The importance of personal norms and situational expectancies to sustainable behaviours: The norm activation and situational expectancy-value theories. Journal of Retailing and Consumer Services, 73, 103371.

Lei, J., & Lin, T. (2022). Emergency online learning: the effects of interactional, motivational, self-regulatory, and situational factors on learning outcomes and continuation intentions. The International Review of Research in Open and Distributed Learning, 23(3), 43-60.

Lei, P., Kong, W., Han, S., Lv, S., & Wang, X. (2022). The Mathematical Culture in Test Items of National College Entrance Examination in China from 1978 to 2021. Mathematics, 10(21), 3987.

Lekhetho, M. (2013). Stakeholder perspectives on strategies that can improve student performance in the school-leaving examinations in Lesotho. Mediterranean Journal of Social Sciences, 4(3), 85.

Lenzi, M., Sharkey, J., Vieno, A., Mayworm, A., Dougherty, D., & Nylund‐Gibson, K. (2015). Adolescent gang involvement: The role of individual, family, peer, and school factors in a multilevel perspective. Aggressive behaviour, 41(4), 386-397.

Leonard, S., Danladi, M., Augustine, A., & Christiana, A. (2022). The Influence of Parental Background on the Academic Performance of Junior Secondary School Students in Ankpa LGA of Kogi State. American Journal of Education and Learning, 7(2), 112-122.

Levpušček, M. P., Zupančič, M., & Sočan, G. (2013). Predicting performance in mathematics in adolescent students: The role of individual and social factors. The Journal of Early Adolescence, 33(4), 523-551.

Li, B.Q., & Wei, X.M. (2017). The Risk of the New College Entrance Examination Reform and its Avoidance--Thinking based on the Risk Society Theory. Educational Development Research, 12, 22–29.

Li, C., & Wang, X. (2015). A Feature Selection Method Based on Competition Winners Mechanism. In 2015 International Power, Electronics and Materials Engineering Conference (pp. 638-643). Atlantis Press.

Li, H., Liu, J., Zhang, D., & Liu, H. (2021). Examining the relationships between cognitive activation, self‐efficacy, socioeconomic status, and achievement in mathematics: A multi‐level analysis. British Journal of Educational Psychology, 91(1), 101-126.

Li, J. T., & Rivera, H. H. (2025). The Relationship Between Teachers' Professional Knowledge and Skills, Teachers' Empowerment, and Students' Learning. TESOL Journal, 16(2), e70024.

Li, J. W., Fei, X., Li, C. Y., Dang, T. Z., & Gao, K. N. (2019). Application of data mining techniques in student performance prediction model. China Education Informatisation(15),55-60. (in Chinese)

Li, J., & He, R. (2022). Family Time and Money Inputs in Education and Teenager Development: Interpretation of Social Capital, Cultural Capital, and Shadow Education. Best Evidence in Chinese Education, 11(1), 1455-1460.

Li, J., Feng, X., Jiang, D., & Zhu, F. (2023, February). Learning Quality Analysis System for the Second Degree Education Based On K-means Algorithm. In 2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (Vol. 6, pp. 1672-1679). IEEE.

Li, J., Li, Y., & You, Y. (2018). A Comparative Study of Curriculum and Classroom Form between Chinese and Japanese Universities: Take Chiba University, Waseda University as an Example. Proceedings of the International Conference on Education, Psychology, and Management Science, 286-290.

Li, L., & Xu, Y. (2023). A study on the relationship between mental toughness and subjective well-being of high school students. BCP Education & Psychology, 9, 229–232.

Li, M., Liu, M., Bin, Y., & Xia, J. (2020). Prediction of circRNA-disease associations based on inductive matrix completion. BMC medical genomics, 13, 1-13.

Li, Q., Yao, Y., & Zhu, X. (2024). The association between writing motivation and performance among primary school students: considering the role of self-efficacy. Humanities and Social Sciences Communications, 11(1), 1-10.

Li, W., & Tong, J. (2023). Impact of the New College Entrance Examination Policy on Students' Access to Higher Education in Shanghai. Journal of Education, Humanities and Social Sciences, 8, 310-318.

Li, Y., Hu, T., Ge, T., & Auden, E. (2019). The relationship between home-based parental involvement, parental educational expectation and academic performance of middle school students in mainland China: A mediation analysis of cognitive ability. International Journal of Educational Research, 97, 139-153.

Liang, X., & Zhang, J. (2023). The Impact of Family Background on Subject Selection in Senior Secondary School Students in the Context of the Reform of College Entrance Examination: An Empirical Study Based on Survey Data from Zhejiang Province. Best Evidence in Chinese Education, 15(1), 1783-1787.

Liang, Z., Zhang, M., Huang, F., Kang, D., & Xu, L. (2021). Application Innovation of Educational Measurement Theory, Method, and Technology in China's New College Entrance Examination Reform. Chinese/English Journal of Educational Measurement and Evaluation|, 2(1), 3.

Liao, C. H., Kuo, B. C., Tsao, C. J., & Mok, M. M. C. (2020). Predictors of Chinese reading and literacy skills among Chinese school children: A 3-year longitudinal study. Educational Psychology, 40(7), 838-855.

Liem，G. (2022). Perceived competence and coping strategies. Educational Psychology, 42(3), 275-277.

Limniou, M., Varga-Atkins, T., Hands, C., & Elshamaa, M. (2021). Learning, Student Digital Capabilities and Academic Performance over the COVID-19 Pandemic. Education Sciences, 11(7), 361.

Lin, B., Zhu, M. S (2013). History of the Chinese Imperial Examination (In Chinese). BEIJING BOOK CO. INC.

Lin, D., Sun, H., & McBride, C. (2019). Morphological awareness predicts the growth rate of Chinese character reading. Developmental science, 22(4), e12793.

Littlejohn, A. (2020). Seeking and sending signals: Remodelling teaching practice during the Covid-19 crisis. ACCESS: Contemporary Issues in Education, 40(1), 56-62.

Liu, C., Wang, H., & Yuan, Z. (2022). A method for predicting the academic performances of college students based on education system data. Mathematics, 10(20), 3737.

Liu, H., Wang, Y., & Wang, H. (2025). Exploring the mediating roles of motivation and boredom in basic psychological needs and behavioural engagement in English learning: a self-determination theory perspective. BMC psychology, 13(1), 179.

Liu, J., & Dai, Z. (2011). The Impact of the Advent of English in Primary Schools on the Development of College English in China. Higher Education Studies, 1(1), 105-112.

Liu, J., Asokan, N., & Pinkas, B. (2015). Secure deduplication of encrypted data without additional independent servers. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (pp. 874-885).

Liu, T. (2021). Orientation and Function of Examination Evaluation in the Reform of College Entrance Examination. Journal of Contemporary Educational Research, 5(8), 215-219.

Liu, X. (2017). The offsite National College Entrance Examination (NCEE) policy in China (Doctoral dissertation, University of Warwick).

Liu, Y., & Zeng, L. (2023, April). Improved Apriori Algorithm in Higher Vocational English Education Information Data Mining. In 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1-5). IEEE.

Liu, Y., Peng, P., & Yan, X. (2025). Early numeracy and mathematics development: A longitudinal meta-analysis on the predictive nature of early numeracy. Journal of Educational Psychology.

Long, Y. (2023). Design and Development of High School English Elective Courses based on the Reform of New College Entrance Examination. Frontiers in Science and Engineering, 3(7), 44–48.

Long, Y., Jia, C., Luo, X., Sun, Y., Zuo, W., Wu, Y., Wu, Y. C, Kaierdebieke, A., & Lin, Z. (2022). The impact of higher education on health literacy: a comparative study between urban and rural China. Sustainability, 14(19), 12142.

Longo, N. V., & Perrotti, C. (2022). Teaching Democratically. A Book Review of Civic Literacy in Schools and Communities. Democracy and Education, 30(1), 10.

Lopes, J., Oliveira, C., & Costa, P. (2022). School and student determinants of reading performance: A multilevel analysis with Portuguese students. Revista de Psicodidáctica (English ed.), 27(1), 29-37.

Lu, H., Leung, F. K., & Fan, Z. (2022). Chinese language and students' mathematics learning: a meta-analysis. ZDM–Mathematics Education, 54(3), 513-528.

Luke, G., & Jude, P. (2025, March). Potluck: Advancing ecological citizenship through" Potluck" teaching: Integrating transdisciplinary learning in mainstream education. In BERA Conference. British Educational Research Association.

Luminoque, V. (2022). Reciprocal Peer-Tutoring on Secondary Students' Achievement in Physics. Psychology, Reciprocal Peer-Tutoring on Secondary Students' Achievement in Physics.

Luo, J., Zhang, K., Ma, W., & Abbasi, M. Z. (2022). The Influence of Family Cultural Capital on Migrant Children's Academic Performance — From the Perspective of Emotional Regulation. International Journal of Neuropsychopharmacology, 25(Supplement_1), A4-A5.

Luo, R., Teng, K., Xu, L., & Zhu, Y. (2021). A decomposition of student achievement gap by gender in China: Evidence from random class assignment. International Journal of Educational Research, 106, 101721.

Lv, T., Tang, P., & Zhang, J. (2023). A Real-Time AIS Data Cleaning and Indicator Analysis Algorithm Based on Stream Computing. Scientific Programming, 2023.

Ma, L., & Zhao, Z. (2025). Reading Motivation and Reading Comprehension Achievement Among English Majors in China: A Descriptive Correlational Study. Heliyon 11 (2025) e42427.

Ma, S., Jia, N., Wei, X., & Zhang, W. (2025). Constructing a predictive model of negative academic emotions in high school students based on machine learning methods. Scientific Reports, 15(1), 1-11.

Ma, Y., Jia, X., Pappas, L., Guo, Y., Feng, T., Feng, J., & Rozelle, S. (2025). The gender gap in early language development among children from peri-urban China. Early Childhood Research Quarterly, 70, 154-166.

Maaliw III, R. R., & Ballera, M. A. (2017). Classification of Learning Styles in Virtual Learning Environment Using J48 Decision Tree. International Association for Development of the Information Society.

Madden, A. D., Webber, S., Ford, N., & Crowder, M. (2018). The relationship between students' subject preferences and their information behaviour. Journal of Documentation, 74(4), 692-721.

Madigan, D. J., & Kim, L. E. (2021). Does teacher burnout affect students? A systematic review of its association with academic performance and student-reported outcomes. International journal of educational research, 105, 101714.

Maghari, A., & Mousa, H. (2017). School students' performance predication using data mining classification. International Journal of Advanced Research in Computer and Communication Engineering, 6(8), 136-141.

MAHALAKSHMI, S., NILAKANTA, S., & BASKAR, G. (2018). Big Data and Parkinson's Disease: Exploration, Analyses, and Data Challenges. In Proceedings of the 51st Hawaii International Conference on Syste, Science (pp. 2778-2783). HICSS.

Malik, A., Onyema, E. M., Dalal, S., Lilhore, U. K., Anand, D., Sharma, A., & Simaiya, S. (2023). Predicting students' adaptability in online entrepreneurship education using modified ensemble machine learning model. Array, 19, 100303.

Malini, J., & Kalpana, Y. (2021). Investigation of factors affecting student performance evaluation using education materials data mining technique. Materials Today: Proceedings, 47, 6105-6110.

Mallika, S., & Mohammed, L. A. (2024). Mathematics through a linguistic lens: The impact of English language proficiency in solving mathematical word problems among secondary ESL students. International Journal of Emerging Issues in Social Science, Arts and Humanities (IJEISSAH), 2(2), 46-57.

Mamaril, J. C. O., & Ballera, M. A. (2022). Multiple educational data mining approaches to discover patterns in university admissions for program prediction. International Journal of Information and Communication Technology, 11(1), 45-56.

Mamat, M. (2016). Penguasaan Bahasa Inggeris dalam kalangan pelajar tahun satu di UKM. Issues in Language Studies, 5(2).

Mamo, H., Gosa, G., & Hailu, B. (2017). Perception of university female students on factors affecting their academic performance and competency: A study from dire Dawa University, Ethiopia. Science Journal of Education, 5(5), 211.

Marsch, L. A., & Gustafson, D. H. (2013). The role of technology in health care innovation: a commentary. Journal of dual diagnosis, 9(1), 101-103.

Marsh, H. W. (1992). Content specificity of relations between academic achievement and academic self-concept. Journal of educational psychology, 84(1), 35.

Martínez-Abad, F., Gamazo, A., & Rodriguez-Conde, M. J. (2020). Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment. Studies in Educational Evaluation, 66, 100875.

Martín-García, A., Martínez-Abad, F., & Reyes-González, D. (2019). TAM and stages of adoption of blended learning in higher education by application of data mining techniques. British Journal of Educational Technology, 50(5), 2384-2400.

Martinot, D., Sicard, A., Gul, B., Yakimova, S., Taillandier-Schmitt, A., & Maintenant, C. (2022). Peers and teachers as the best source of social support for school engagement for both advantaged and priority education area students. Frontiers in Psychology, 13, 958286.

Mastrantonio, M. P., Pestana, J. V., & Codina, N. C. M. (2023). Predicting procrastination with academic performance: Towards the anticipation of a higher education problem. Intangible Capital, 2023, vol. 19, num. 2, p. 316-327.

Materu, G., Robinson, N., & Joshua, J. M. (2025). Evaluating the alignment of classroom practices with stated instructional objectives in selected secondary schools in Morogoro municipality. International Journal of Education and Social Science Research, 8(2), 412-437.

Mathimagal, N., Jayalakshmi, S., & Prasanna, S. (2022, May). Impact of Social Media Affecting Student's Academic Performance using Opinion Mining. In 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 722-729). IEEE.

Matzavela, V., & Alepis, E. (2021). Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments. Computers and Education: Artificial Intelligence, 2, 100035.

Mayaza, S. B., Buaton, R., & Ramadani, S. (2023). Application of Data Mining in Analyzing the Effect of Parents' Employment and Education Level on Student Behaviour Using the A PRIORI Method (Case Study: SDN 024769 Binjai). Journal of Artificial Intelligence and Engineering Applications (JAIEA), 3(1), 164-171.

Mbogo, R. W., & Wekesa, F. C. (2021). The influence of social economic background on academic performance of secondary school student council officials in Kenya. Editon Consortium Journal of Curriculum and Educational Studies, 3(1), 331-339.

McClung, N. A., & Arya, D. J. (2018, May). Individual differences in fourth-grade math performance in Chinese and English. In Frontiers in Education (Vol. 3, p. 29). Frontiers Media SA.

McGuire, A. L. (2008). Identifiability of DNA data: the need for consistent federal policy. The American journal of bioethics: AJOB, 8(10), 75.

Meaders, C. L., Mendez, L., Aguilar, A. G., Rivera, A. T., Vasquez, I., Mueller, L. O., & Owens, M. T. (2025). An Asynchronous Chemistry-in-biology Intervention Improves Student Content Knowledge and Performance in Introductory Biology. CBE—Life Sciences Education, 24(1), ar2.

Mehndiratta, A., & Mehndiratta, P. (2023, August). Factors Affecting Student's Academic Performance In Programming Using Association Rule Mining. In Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing (pp. 371-374).

Mendez, L., Rivera, A. T., Vasquez, I., Godínez Aguilar, A., Owens, M. T., & Meaders, C. L. (2024). How students taking introductory biology experience the chemistry content. Journal of Microbiology and Biology Education, 25(3), e00111-24.

Menter, Z., Tee, W. Z., & Dave, R. (2021). Application of machine learning-based pattern recognition in iot devices. In Proceedings of International Conference on Communication and Computational Technologies: ICCCT 2021 (pp. 669-689). Springer Singapore.

Merema, B., Breesch, H., & Saelens, D. (2019). Comparison of model identification techniques for MPC in all-air HVAC systems in an educational building. In E3S Web of Conferences (Vol. 111, p. 01053). EDP Sciences.

Meriläinen, N., & Hiljanen, M. (2024). Power relations in society. Frontiers in Political Science, 6, 1477676.

Mi, H., Gao, Z., Zhang, Q., & Zheng, Y. (2022). Research on Constructing Online Learning Performance Prediction Model Combining Feature Selection and Neural Network. International Journal of Emerging Technologies in Learning, 17(7)，94.

Miao, J., Tian, X., Zhang, L., Zhang, B., & Ma, L. (2025). An analysis of the empirical research on SL/FL teaching and learning based on positive psychology: a systematic literature review. BMC psychology, 13(1), 1-12.

Michel, P., Ngo, N., Pons, J. F., Delliaux, S., & Giorgi, R. (2021). A filter approach for feature selection in classification: application to automatic atrial fibrillation detection in electrocardiogram recordings. BMC medical informatics and decision making, 21(4), 1-17.

Mike, K., & Hazzan, O. (2022, March). Interdisciplinary CS1 Course for Non-Majors: The Case of Graduate Psychology Students. In 2022 IEEE global engineering education conference (EDUCON) (pp. 86-93). IEEE.

Miksza, P., Evans, P., & McPherson, G. E. (2021). Wellness among university-level music students: A study of the predictors of subjective vitality. Musicae Scientiae, 25(2), 143-160.

Miller, D., & Trainin, G. (2023). Impact of the Pandemic on Computer Science Education. Research and Evaluation in Education, Technology, Art, and Design, 46.

Mishra, S. (2020). Social networks, social capital, social support and academic success in higher education: A systematic review with a special focus on 'under represented' students. Educational Research Review, 29, 100307.

Mittal, S., Mahendra, S., Sanap, V., & Churi, P. (2022). How can machine learning be used in stress management: A systematic literature review of applications in workplaces and education. International Journal of Information Management Data Insights, 2(2), 100110.

Mkwazu, H. R., & Yan, C. (2020, October). Grade prediction method for university course selection based on decision tree. In Proceedings of the 2020 International Conference on Aviation Safety and Information Technology (pp. 593-599).

Mohamed, A. A., Dahie, A. M., & Warsame, A. A. (2018). Factors affecting student academic performance: Case study from University of Somalia in Mogadishu-Somalia. IOSR Journal of Humanities and Social Science (IOSR-JHSS), 23(3), 73-10.

Mohammed, S., Elbeltagi, A., Bashir, B., Alsafadi, K., Alsilibe, F., Alsalman, A., Zeraatpisheh, M., Széles, A., & Harsányi, E. (2022). A comparative analysis of data mining techniques for agricultural and hydrological drought prediction in the eastern Mediterranean. Computers and Electronics in Agriculture, 197, 106925.

Moldabayev, D., Menicucci, J., Al-Zubaidy, S., & Abdulaziz, N. (2013). Attendance, performance and culture Experience of the School of Engineering, Nazarbayev University - An update. In 2013 IEEE Global Engineering Education Conference, EDUCON 2013 (pp. 5-10).

Molina Huerta, C., Sotelo Atahua, A., Villacrisis Guerrero, J., & Andrade-Arenas, L. (2023). Data mining: Application of digital marketing in education. Advances in Mobile Learning Educational Research, 3(1), 621-629.

Moliner, L., & Alegre, F. (2020). Effects of peer tutoring on middle school students' mathematics self-concepts. Plos one, 15(4), e0231410.

Moliner, L., & Alegre, F. (2022). Peer tutoring in middle school mathematics: Academic and psychological effects and moderators. Educational Psychology, 42(8), 1027-1044.

Montague, J. R. (1995). Exam performance and GPA for first-semester biology students. Journal of College Science Teaching, 24(4), 245-248.

Moreira, G. D. O., Passeri, S., Velho, P. E. N. F., Ferraresi, F., Appenzeller, S., & Amaral, E. (2019). The academic performance of scholarship students during medical school. Revista Brasileira de Educação Médica, 43, 163-169.

Munda, X., & Tiwari, V. K. (2024). The Impact of Academic Procrastination on Students' Performance in Indian School Education Systems: A Special Research Analysis-Vision 2045.

Munir, H., Vogel, B., & Jacobsson, A. (2022). Artificial intelligence and machine learning approaches in digital education: a systematic revision. Information, 13(4), 203.

Murphy, P. (2022, July). The Home Exam: from Reading List Reproduction to Student Centred Production. In Conference Proceedings. The Future of Education 2022.

Musa, A. B. (2024). Understanding Student Performance in Foundation Year: Insights from Logistic Regression, Naïve Bayes, and Random Forest Models. International Journal of Information and Education Technology, 14(12).

Musgrove, M. M. C., Ko, M. E., Schinske, J. N., & Corwin, L. A. (2024). Broadening participation in biology education research: A role for affinity groups in promoting social connectivity, self-efficacy, and belonging. CBE—Life Sciences Education, 23(1), 1-17.

Mustafa, M., Alshare, M., Bhargava, D., Neware, R., Singh, B., & Ngulube, P. (2022). Perceived Security Risk Based on Moderating Factors for Blockchain Technology Applications in Cloud Storage to Achieve Secure Healthcare Systems. Computational and Mathematical Methods in Medicine, 2022, 6112815-6112815.

Muthalagu, A., Pacheco, J. A., Aufox, S., Peissig, P. L., Fuehrer, J. T., Tromp, G., Kho, A. N., & Rasmussen-Torvik, L. J. (2014). A rigorous algorithm to detect and clean inaccurate adult height records within EHR systems. Applied Clinical Informatics, 5(01), 118-126.

Naim, A. (2022). E-learning engagement through convolution neural networks in business education. European Journal of Innovation in Nonformal Education, 2(2), 497-501.

Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. Review of educational research, 89(3), 459-496.

Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237.

Nasution, M. K. (2022). Social network mining (SNM): A definition of relation between the resources and SNA. arXiv preprint arXiv:2207.06234.

Nasution, M. Z. F., Sitompul, O. S., & Ramli, M. (2018). PCA based feature reduction to improve the accuracy of decision tree c4. 5 classification. In Journal of Physics: Conference Series (Vol. 978, No. 1, p. 012058). IOP Publishing.

Nazeef, N. M., & Ali, J. (2024). Impact of Collaborative Learning on Students Academic Performance in Teacher's Education Program. Journal of Asian Development Studies, 13(1), 1054-1068.

Neville-Barton, P., & Barton, B. (2005). The relationship between English language and mathematics learning for non-native speakers.

Ngito, J. A., & Kwaba, J. G. (2022). Influence of Parent's Level of Income on Student's Enrolment in Secondary Schools in Rachuonyo East Sub-County, Homabay County, Kenya. East African Journal of Education Studies, 5(3), 17-28.

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. Education and Information Technologies, 28(4), 4221-4241.

Nguyen-Huy, T., Deo, R. C., Khan, S., Devi, A., Adeyinka, A. A., Apan, A. A., & Yaseen, Z. M. (2022). Student performance predictions for advanced engineering mathematics course with new multivariate copula models. IEEE Access, 10, 45112-45136.

Ni, A., Cheung, A. C., & Shi, J. (2022). Effects of educational technology on reading performance for Chinese K-12 English second language learners: A meta-analysis. Frontiers in Psychology, 13, 1025761.

Niazkar, H. R., & Niazkar, M. (2020). Application of artificial neural networks to predict the COVID-19 outbreak. Global Health Research and Policy, 5(1), 1-11.

Niehaus, K., & Adelson, J. L. (2014). School support, parental involvement, and academic and social-emotional outcomes for English language learners. American Educational Research Journal, 51(4), 810-844.

Nilsen, T., & Gustafsson, J. E. (2016). The impact of school climate and teacher quality on mathematics performance: A difference-in-differences approach. Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time, 2, 81-95. Springer.

Niu, J., Xu, H., & Yu, J. (2025). Identifying multilevel factors on student mathematics performance for Singapore, Korea, Finland, and Denmark in PISA 2022: considering individualistic versus collectivistic cultures. Humanities and Social Sciences Communications, 12(1), 1-12.

Nnoli, J. N. (2024). The efficacy of teacher's qualification and teaching experience in academic achievement of chemistry students. Pakistan Journal of Educational Research, 7(3), 1-12.

Nuarini, S., Fauziah, S., Mayangky, N. A., & Nurfalah, R. (2023). Comparison Algorithm on Machine Learning for Student Mental Health Data. Journal Medical Informatics Technology, 81-85.

Nyamubi, G. J. (2019). Socio-economic status as determinants of students' performance in English language in secondary schools in Tanzania. Education journal, 8(3), 110-119.

O'Connor, L. M., O'Connor, B. A., Lim, S. B., Zeng, J., & Lo, C. H. (2023). Integrative multi-omics and systems bioinformatics in translational neuroscience: A data mining perspective. Journal of Pharmaceutical Analysis, 13(8), 836-850.

Ocaña-Fernández, Y. J., Gómez-Gonzales, W., Fernández, L. A. V., Ramos, S. P. V., Zúñiga, H. F. D., Fernandez, J. R. H., & Broncano, M. A. A. (2023). A Novel Approach to Predict the Early Childhood Special Education Learning Skills of Autistic Children Using Ensemble Machine Learning. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), 14(2), 59-65.

Okano, K., Kaczmarzyk, J. R., Dave, N., Gabrieli, J. D., & Grossman, J. C. (2019). Sleep quality, duration, and consistency are associated with better academic performance in college students. NPJ science of learning, 4(1), 16.

Okereke, G. E. (2020). A machine learning based framework for predicting student's academic performance. Phys Sci Biophys J, 4(2), 000145.

Okewu, E., Adewole, P., Misra, S., Maskeliunas, R., & Damasevicius, R. (2021). Artificial neural networks for educational data mining in higher education: A systematic literature review. Applied Artificial Intelligence, 35(13), 983-1021.

Olawale, B. E., & Hendricks, W. (2024). Mathematics teachers' self-efficacy beliefs and its relationship with teaching practices. Eurasia Journal of Mathematics, Science and Technology Education, 20(1), em2392.

Oredein, A., & Awodun, A. (2013). Impact of Teachers' Motivational Indices on Science Students' Academic Performance in Nigerian Senior Secondary Schools. International Education Studies, 6(2), 49-54.

Ouyang, Y. (2021). Study on the Continuation Task of China's New College Entrance Examination Based on the Interactive Alignment Theory. International Journal of Linguistics, Literature and Translation, 4(10), 89-91.

Owusu-Boadu, B., Nti, I. K., Nyarko-Boateng, O., Aning, J., & Boafo, V. (2021). Academic performance modelling with machine learning based on cognitive and non-cognitive features. Applied Computer Systems, 26(2), 122-131.

Ozdemir, N. (2019). A bibliometric analysis of research on principal leadership and student performance. New Trends and Issues Proceedings on Humanities and Social Sciences, 6(7), 55-61.

Pahuriray, A. V., Basanta, J. D., Arroyo, J. C. T., & Delima, A. J. P. (2022). Flexible Learning Experience Analyser (FLExA): Sentiment Analysis of College Students through Machine Learning Algorithms with Comparative Analysis using WEKA. Int. J. Emerg. Technol. Adv. Eng, 12, 1-15.

Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. Entropy, 23(4), 485.

Panayiotou, M., Humphrey, N., & Wigelsworth, M. (2019). An empirical basis for linking social and emotional learning to academic performance. Contemporary Educational Psychology, 56, 193-204.

Pandey, M., & Sharma, V. K. (2013). A decision tree algorithm pertaining to the student performance analysis and prediction. International Journal of Computer Applications, 61(13), 1-5.

Pang, W., Xu, X., Lin, L., & Ren, Y. (2013). The impact of family socioeconomic status on students' academic achievement. Global Education, 42(2), 12-21.

Pant, M. M. P., & Chowhan, S. (2023). Identification and comparison of frequent steps for evaluating and designing online admission systems using data mining in R and Weka referring to professional programmes. The Online Journal of Distance Education and e-Learning, 11(2), 1407-1416.

Park, E., & Yi, S. (2025). Compulsory education enhances financial inclusion across socioeconomic groups: a global analysis. Humanities and Social Sciences Communications, 12(1), 1-14.

Park, Y. E. (2020). Uncovering trend-based research insights on teaching and learning in big data. Journal of Big Data, 7(1), 93.

Parmar, P., & Nathans, L. (2022). Parental Warmth and Parent Involvement: Their Relationships to Academic Performance and Behaviour Problems in School and Related Gender Effects. Societies, 12(6), 1-20.

Parr, A. K., & Bonitz, V. S. (2015). Role of family background, student behaviours, and school-related beliefs in predicting high school dropout. The Journal of Educational Research, 108(6), 504-514.

Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2020). The impact of stress on students in secondary school and higher education. International journal of adolescence and youth, 25(1), 104-112.

Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. International Journal of Computer Sciences and Engineering, 6(10), 74-78.

Paudyal, A. P. B. S., Ramsey, S., & Dunlap, K. (2023). 189 Network Analysis of Survey Results from a Large Enrolment Course: Investigating Dynamics Among Background, Career Aspirations and Decision-Making of Students Related to Anima Science. Journal of Animal Science, 101(Supplement_1), 114-115.

Pawlina, W. (2019). Not “how should I learn?” or “how should I act?” but,“who shall I become?”: A précis on the roots of early professional identity formation in the anatomy course. Anatomical Sciences Education, 12(5), 465-467.

Peixoto, F., Radišić, J., Krstić, K., Hansen, K. Y., Laine, A., Baucal, A., & Mata, L. (2023). Contribution to the Validation of the Expectancy-Value Scale for Primary School Students. Journal of Psychoeducational Assessment, 41(3), 343-350.

Peng, P., & Kievit, R. A. (2020). The development of academic performance and cognitive abilities: A bidirectional perspective. Child Development Perspectives, 14(1), 15-20.

Peng, P., Lin, X., Ünal, Z. E., Lee, K., Namkung, J., Chow, J., & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. Psychological Bulletin, 146(7), 595.

Perez, J. G., & Perez, E. S. (2021). Predicting student program completion using Naïve Bayes classification algorithm. International Journal of Modern Education and Computer Science, 13(3), 57-67.

Phauk, S., & Okazaki, T. (2021). Integration of Educational Data Mining Models to a Web-Based Support System for Predicting High School Student Performance. International Journal of Computer and Information Engineering, 15(2), 131-144.

Porter, S. B., Odegard, T. N., Farris, E. A., & Oslund, E. L. (2024). Effects of teacher knowledge of early reading on students' gains in reading foundational skills and comprehension. Reading and Writing, 37(8), 2007-2023.

Pradana, D. A., Mahfud, M., Hermawan, C., & Susanti, H. D. (2020). Nasionalism: Character education orientation in learning development. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Volume, 3, 4026-4034.

Pramudi, G. F., Firmansyah, G., Tjahjono, B., & Widodo, A. M. (2023). Analysis of School Community Sentiment towards Personal Data Protection Law Using Support Vector Machine (SVM) Method. Asian Journal of Social and Humanities, 1(12), 1256-1275.

Prince, D., Madhan, K., Vishwa, K., & Yamunathangam, D. (2023, June). Job and Course Recommendation System using Collaborative Filtering and Naive Bayes algorithms. In 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) (pp. 1-4). IEEE.

Pritchard, M. E., & Wilson, G. S. (2003). Using emotional and social factors to predict student success. Journal of college student development, 44(1), 18-28.

Priyanka, & Kumar, D. (2020). Decision tree classifier: a detailed survey. International Journal of Information and Decision Sciences, 12(3), 246-269.

Pulfrey, C., Darnon, C., & Butera, F. (2013). Autonomy and task performance: Explaining the impact of grades on intrinsic motivation. Journal of educational psychology, 105(1), 39-57.

Pulgar, J., Ramírez, D., Umanzor, A., Candia, C., & Sánchez, I. (2022). Long-term collaboration with strong friendship ties improves academic performance in remote and hybrid teaching modalities in high school physics. Physical Review Physics Education Research, 18(1), 010146.

Pumptow, M., & Brahm, T. (2021). Students' digital media self-efficacy and its importance for higher education institutions: development and validation of a survey instrument. Technology, Knowledge and Learning, 26, 555-575.

Putwain, D. W., Daly, A. L., Chamberlain, S., & Sadreddini, S. (2016). 'Sink or swim': buoyancy and coping in the cognitive test anxiety–academic performance relationship. Educational Psychology, 36(10), 1807-1825.

Putwain, D. W., von der Embse, N. P., Rainbird, E. C., & West, G. (2020). The development and validation of a new Multidimensional Test Anxiety Scale (MTAS). European Journal of Psychological Assessment.

Putwain, D., & Daly, A. L. (2014). Test anxiety prevalence and gender differences in a sample of English secondary school students. Educational Studies, 40(5), 554-570.

Qorie, T., Ahmadi, A., & Atiqoh, A. (2024). Classroom Activities of Cooperative Learning in Chinese as a Foreign Language. Sinolingua: Journal of Chinese Studies, 2(1), 21-29.

Rabiu, H., Muhammed, A. I., Umaru, Y., & Ahmed, H. T. (2016). Impact of mobile phone usage on academic performance among secondary school students in Taraba State, Nigeria. European scientific journal, 12(1), 466-479.

Rafiola, R., Setyosari, P., Radjah, C., & Ramli, M. (2020). The effect of learning motivation, self-efficacy, and blended learning on students' performance in the industrial revolution 4.0. International Journal of Emerging Technologies in Learning (iJET), 15(8), 71-82.

Ragab, M., Abdel Aal, A. M. K., Jifri, A. O., & Omran, N. F. (2021). Enhancement of Predicting Students Performance Model Using Ensemble Approaches and Educational

Data Mining Techniques. Wireless Communications and Mobile Computing, 2021, 1-12.

Rahman, M. (2021). Impact of socio-economic factors on undergraduate students' academic performance in Bangladesh: a case study at social science faculty, University of Dhaka. Asian J. Soc. Sci. Leg. Stud, 3(4), 147-157.

Rahman, M. M., Watanobe, Y., Matsumoto, T., Kiran, R. U., & Nakamura, K. (2022). Educational data mining to support programming learning using problem-solving data. IEEE Access, 10, 26186-26202.

Rajagukguk, S. A., Fudholi, D. H., & Pratama, A. R. (2023). A Data-Driven Causal Modelling Analysis of Socio-Economic Factors and Its Impact on Student's Performance: A Case Study of a Junior High School in Bali. AL-ISHLAH: Jurnal Pendidikan, 15(2), 2481-2497.

Rao, K. S., Swapna, N., & Kumar, P. (2017). Educational data mining for student placement prediction using machine learning algorithms. International Journal of Engineering & Technology, 7(1.2), 44-47.

Rashid, A., Asif, S., Anwer Butt, N., & Ashraf, I. (2013). Feature Level Opinion Mining of Educational Student Feedback Data using Sequential Pattern Mining and Association Rule Mining. International Journal of Computer Applications, 81(10), 31-38.

Rawal, A., & Lal, B. (2023). Predictive model for admission uncertainty in high education using Naïve Bayes classifier. Journal of Indian Business Research, 15(2), 262-277.

Razaque, A., Abenova, M., Alotaibi, M., Alotaibi, B., Alshammari, H., Hariri, S., & Alotaibi, A. (2022). Anomaly detection paradigm for multivariate time series data mining for healthcare. Applied Sciences, 12(17), 8902.

Remali, N. A. S., Shamsuddin, M. R., & Abdul-Rahman, S. (2022, September). Sentiment Analysis on Online Learning for Higher Education During Covid-19. In 2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 142-147). IEEE.

Ren, G., & Yang, Y. (2020).Application of BP neural network in early warning of students' performance in colleges and universities. Digital Technology and Application (10), 53-55. (in Chinese) doi:10.19695/j.cnki.cn12-1369.2020.10.18.

Ren, Y., Pu, J., Yang, Z., Xu, J., Li, G., Pu, X., Yu, P. S., & He, L. (2022). Deep clustering: A comprehensive survey. arXiv preprint arXiv:2210.04142.

Rienties, B., Køhler Simonsen, H., & Herodotou, C. (2020). Defining the boundaries between artificial intelligence in education, computer-supported collaborative learning, educational data mining, and learning analytics: A need for coherence. In frontiers in Education (Vol. 5, p. 128). Frontiers Media SA.

Rivera Rodas, E. I. (2019). Separate and Unequal--Title I and Teacher Quality. Education policy analysis archives, 27(14), n14.

Robu, R., Arseni-Ailoi, P., & Ungureanu-Anghel, D. (2023, May). Using Weka API for creating a custom classification application. In 2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 000535-000538). IEEE.

Rodríguez-Arribas, S., Marticorena-Sánchez, R., & Ji, Y. (2021). Active methodology, educational data mining and learning analytics: A systematic mapping study. Informatics in Education, 20(2), 171-204. https://doi.org/10.15388/infedu.2021.09

Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data mining and knowledge discovery, 3(1), 12-27.

Rosa, E. M., & Tudge, J. (2013). Urie Bronfenbrenner's theory of human development: Its evolution from ecology to bioecology. Journal of Family Theory & Review, 5(4), 243–258.

Roslan, M. H. B., & Chen, C. J. (2023). Predicting students' performance in English and Mathematics using data mining techniques. Education and Information Technologies, 28(2), 1427-1453.

Rost, D. H., & Feng, X. (2024). Academic self-concept wins the race: The prediction of achievements in three major school subjects by five subject-specific self-related variables. Behavioural Sciences, 14(1), 40.

Rozek, C. S., Ramirez, G., Fine, R. D., & Beilock, S. L. (2019). Reducing socioeconomic disparities in the STEM pipeline through student emotion regulation. Proceedings of the National Academy of Sciences, 116(5), 1553-1558.

Rubach, C., Dicke, A. L., Safavian, N., & Eccles, J. S. (2023). Classroom transmission processes between teacher support, interest value and negative affect: An investigation guided by situated expectancy-value theory and control-value theory. Motivation and Emotion, 1-20.

Saeed, K. M., Ahmed, A. S., Rahman, Z. M., & Sleman, N. A. (2023). How social support predicts academic achievement among secondary students with special needs: the mediating role of self-esteem. Middle East Current Psychiatry, 30(1), 46.

Sağkal, A. S., & Sönmez, M. T. (2022). The effects of perceived parental math support on middle school students' math engagement: the serial multiple mediation of math self-efficacy and math enjoyment. European Journal of Psychology of Education, 37(2), 341-354.

Sahni, J. (2023). Is Learning Analytics the Future of Online Education?: Assessing Student Engagement and Academic Performance in the Online Learning Environment. International Journal of Emerging Technologies in Learning (Online), 18(2), 33.

Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., & Fischer, F. (2023). Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. Learning and Instruction, 83, 101620.

Sáiz-Manzanares, M. C., Marticorena-Sánchez, R., & Ochoa-Orihuel, J. (2021). Using advanced learning technologies with university students: an analysis with machine learning techniques. Electronics, 10(21), 2620.

Sáiz-Manzanares, M. C., Rodríguez-Díez, J. J., Díez-Pastor, J. F., Rodríguez-Arribas, S., Marticorena-Sánchez, R., & Ji, Y. P. (2021). Monitoring of student learning in learning management systems: An application of educational data mining techniques. Applied Sciences, 11(6), 2677.

Salam, M. A., Azar, A. T., Elgendy, M. S., & Fouad, K. M. (2021). The effect of different dimensionality reduction techniques on machine learning overfitting problem. Int. J. Adv. Comput. Sci. Appl, 12(4), 641-655.

Saleh, A., Dharshinni, N. P., Perangin-Angin, D., Azmi, F., & Sarif, M. I. (2023). Implementation of Recommendation Systems in Determining Learning Strategies Using the Naïve Bayes Classifier Algorithm. Sinkron: jurnal dan penelitian teknik informatika, 8(1), 256-267.

Salihoun, M. (2020). State of art of data mining and learning analytics tools in higher education. International Journal of Emerging Technologies in Learning (iJET), 15(21), 58-76.

Sánchez-Hernando, B., Antón-Solanas, I., Júarez-Vela, R., Gea-Caballero, V., Carboneres-Tafaner, M. I., Ferrer-Gracia, E., Gállego-Diéguez, J., Santolalla-Arnedo, I., & Gasch-Gallén, Á. (2021). Healthy Lifestyle and Academic Performance in Middle School Students from the Region of Aragón (Spain). International Journal of Environmental Research and Public Health, 18(16):8624.

Sanchez-Martin, J., Alvarez-Gragera, G. J., Davila-Acedo, M. A., & Mellado, V. (2017). Teaching technology: From knowing to feeling enhancing emotional and content acquisition performance through Gardner's Multiple Intelligences Theory in technology and design lessons. Journal of Technology and Science Education, 7(1), 58-79.

Sang, Z. (2024). Research on the language value of college students' ideological and political education in the Internet era. Applied Mathematics and Nonlinear Sciences, 9(1), 1–12. https://doi.org/10.2478/amns-2024-0814

Santos-Rufo, A., Mesas-Carrascosa, F. J., García-Ferrer, A., & Meroño-Larriva, J. E. (2020). Wavelength selection method based on partial least square from hyperspectral unmanned aerial vehicle orthomosaic of irrigated olive orchards. Remote Sensing, 12(20), 3426.

Saqib, M., & Rehman, K. U. (2018). Impact of stress on students' academic performance at secondary school level at District Vehari. International Journal of Learning and Development, 8(1), 84-93.

Sarwat, S., Ullah, N., Sadiq, S., Saleem, R., Umer, M., Eshmawi, A. A., Mohamed, A., & Ashraf, I. (2022). Predicting Students' Academic Performance with Conditional Generative Adversarial Network and Deep SVM. Sensors, 22(13), 4834.

Sawatzky, R., Ratner, P. A., Johnson, J. L., Kopec, J. A., & Zumbo, B. D. (2009). Sample heterogeneity and the measurement structure of the Multidimensional Students' Life Satisfaction Scale. Social Indicators Research, 94, 273-296.

Sawtelle, V., Brewe, E., & Kramer, L. H. (2012). Exploring the relationship between self-efficacy and retention in introductory physics. Journal of research in science teaching, 49(9), 1096-1121.

Schenke, W., van Driel, J. H., Geijsel, F. P., & Volman, M. L. (2017). Closing the feedback loop: a productive interplay between practice-based research and school development through cross-professional collaboration in secondary education. Professional Development in Education, 43(5), 860-880.

Schlendorf, C., Kelly, A. M., & Krakehl, R. (2025). School-level Earth science course enrolment, student demographics, and Earth science performance. Journal of Geoscience Education, 1-11.

Schneider, R., & Wolff, F. (2023). The formation of subject-specific values as a two-step process: Self-concepts mediate the relation between achievement and values. Contemporary Educational Psychology, 75, 102223.

Schultze, A., Bates, C., Cockburn, J., MacKenna, B., Nightingale, E., Curtis, H. J., Hulme., W. J., & Goldacre, B. (2021). Identifying Care Home Residents in Electronic Health Records-An OpenSAFELY Short Data Report. Wellcome Open Research, 6(90), 90.

Schwartz, A. E., & Rothbart, M. W. (2020). Let them eat lunch: The impact of universal free meals on student performance. Journal of Policy Analysis and Management, 39(2), 376-410.

Setiabudi, D. H., & Santoso, M. (2023). Effect of Students' Activities on Academic Performance Using Clustering Evolution Analysis. CommIT (Communication and Information Technology) Journal, 17(2), 209-219.

Sewell, J. L., Bowen, J. L., Cate, O. T., O'Sullivan, P. S., Shah, B., & Boscardin, C. K. (2020). Learning challenges, teaching strategies, and cognitive load: insights from the experience of seasoned endoscopy teachers. Academic Medicine, 95(5), 794-802.

Seyyedyan, M., Moosazadeh, T., & Begloo, M. N. M. (2021). Comparison of personality trait and emotional intelligence between the women who experienced domestic violence and control group. Journal of Education Culture and Society, 12(1), 212-223.

Sha, T. (2019, May). Analysis of the influence of China's new college entrance examination reform on after-school training industry and educational informatization. In 2019 5th International Conference on Humanities and Social Science Research (ICHSSR 2019) (pp. 271-276). Atlantis Press.

Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2022). Student retention using educational data mining and predictive analytics: a systematic literature review. IEEE Access, 10, 72480-72503.

Shang, C., Moss, A. C., & Chen, A. (2023). The expectancy-value theory: A meta-analysis of its application in physical education. Journal of Sport and Health Science, 12(1), 52-64.

Shantakumar, S. R., Sahabdeen, H. B., Abidin, F. A. B. Z., Perumal, G., & Kumar, N. (2022). Association of type and duration of exercise with the mental and physical health and academic performance of Medical undergraduate students-Cross-sectional study. Bangladesh Journal of Medical Science, 21(1), 135-139.

Shao, Y., Kang, S., Lu, Q., Zhang, C., & Li, R. (2024). How peer relationships affect academic achievement among junior high school students: The chain mediating roles of learning motivation and learning engagement. BMC psychology, 12(1), 278.

Sharma, N., Sharma, M., & Garg, U. (2023, January). Predicting Academic Performance of Students Using Machine Learning Models. In 2023 International Conference on Artificial Intelligence and Smart Communication (AISC) (pp. 1058-1063). IEEE.

Sharma, S., & Gupta, B. (2023). Investigating the role of technostress, cognitive appraisal and coping strategies on students' learning performance in higher education: a multidimensional transactional theory of stress approach. Information Technology & People, 36(2), 626-660.

Shen, C. (2018). Research on students' test-choosing behaviour under the background of new college entrance examination. Zhejiang: Ningbo University, pp. 27–29.

Shi, C. (2025). A Case Study of National Matriculation English Test Validity in Predicting Students' English Performance in College. SAGE Open, 15(2), 21582440251331196.

Simpfendoerfer, L. N., Sundstrom, M., Dew, M., & Holmes, N. G. (2024). What topics of peer interactions correlate with student performance in physics courses?. European Journal of Physics, 45(3), 035704.

Singh, H. P., & Alhulail, H. N. (2022). Predicting Student-Teachers Dropout Risk and Early Identification: A Four-Step Logistic Regression Approach. IEEE Access, 10, 6470-6482.

Song, J., Mullick, J., & Jiang, T. (2025). English language education in Chinese primary schools: exploring EFL teachers' attitudes, self-efficacy, and perceived school environment. Asian-Pacific Journal of Second and Foreign Language Education, 10(1), 18.

Song, Y., S. & Zhang, C., H. (2019). The New Round of College Entrance Examination Reform in China. International Journal of Educational Administration, 11(1), 23-32.

Sorhagen, N. S. (2013). Early teacher expectations disproportionately affect poor children's high school performance. Journal of educational psychology, 105(2), 465.

Staniszewska, S., Hill, E. M., Grant, R., Grove, P., Porter, J., Shiri, T., Tulip, S., Whitehurst, J., Wright, C., Datta, S., Petrou, S., & Keeling, M. (2021). Developing a framework for public involvement in mathematical and economic modelling: bringing new dynamism to vaccination policy recommendations. The Patient-Patient-Centered Outcomes Research, 14, 435-445.

Stark, J. T. (2024). Enhancing Algorithmic Early Warning Systems With Dynamic Selection to Predict High School Graduation Outcomes (Doctoral dissertation, University of Nevada, Reno).

Steinmayr, R., Weidinger, A. F., Schwinger, M., & Spinath, B. (2019). The importance of students' motivation for their academic performance–replicating and extending previous findings. Frontiers in psychology, 10, 1730.

Stewart, E. B. (2008). School structural characteristics, student effort, peer associations, and parental involvement: The influence of school-and individual-level factors on academic performance. Education and urban society, 40(2), 179-204.

Stover, J. B., de la Iglesia, G., Boubeta, A. R., & Liporace, M. F. (2012). Academic Motivation Scale: adaptation and psychometric analyses for high school and college students. Psychology research and behaviour management, 71-83.

Street, K. E. S., Malmberg, L. E., & Schukajlow, S. (2024). Students' mathematics self-efficacy: a scoping review. ZDM–Mathematics Education, 56(2), 265-280.

Stresau, K. S., & Steiner, M. W. (2020). A Framework for Developing a Deeper Understanding of the Factors that Influence Success and Failure in Undergraduate Engineering Capstone Design Experiences. In 2020 ASEE Virtual Annual Conference Content Access.

Sudadio, M., Nurullah, M., Apriani, R., Anwar, U. S., & Karim, A (2024b). Contribution of Education Costs in Improving the Quality of SMP Examination Results Banten Province in 2023.

Sudadio, M., Satiri, Y. H., & Ahmadi, V. M. (2024a). The Contribution of the Principal's Managerial Function in Improving the Quality of Education Based on School Exam Results in High Schools in Banten Province.

Sultan Alalawi, S. J., Mohd Shaharanee, I. N., & Mohd Jamil, J. (2023). Clustering Student Performance Data Using k-Means Algorithms. Journal of Computational Innovation and Analytics (JCIA), 2(1), 41-55.

Sun, J. (2023). The Gaokao and Its History. In Effectiveness and Fairness of Chinese Higher Education Admissions Policy: Perceptions and Reforms (pp. 1-23). Singapore: Springer Nature Singapore.

Sun, X., & Liu, Y. (2019). Key Factors Analysis of Cultivating Research-oriented Talents in China's Universities Based on the Talent Demand for China's Strategic Emerging Industry. 2018 3rd International Conference on Education, Management and Systems Engineering (EMSE 2018).

Sun, Y., Hong, J. C., Ye, J. H., & Ye, J. N. (2023). Satisfaction with Online Study Abroad Predicted by Motivation and Self-Efficacy: A Perspective Based on the Situated Expectancy–Value Theory during the COVID-19 Epidemic. Sustainability, 15(5), 4070.

Sun, Z., & Wu, Q. (2023). Two-phase tabu search algorithm for solving Chinese high school timetabling problems under the new college entrance examination reform. Data Science and Management, 6(1), 55-63.

Suningsih, N. (2022). Examination of the Role of Family Socioeconomic Status and Parental Education in Predicting English as a Foreign Language Learners' Receptive Skills Performance. Journal of English as A Foreign Language Teaching and Research, 2(2), 21-29.

Supervía, U. P., Bordás, S. C., & Robres, Q. A. (2022). The mediating role of self-efficacy in the relationship between resilience and academic performance in adolescence. Learning and Motivation, 78, 101814.

Supovitz, J., & Comstock, M. (2023). The impact of a formal teacher leadership program on student performance. Archivos Analíticos de Políticas Educativas= Education Policy Analysis Archives, 31(1), 74.

Swamy, R. S., Sahabdeen, H. B., Abidin, Z. A. F., Perumal, G., & Kumar, N. (2022). Association of type and duration of exercise with the mental and physical health and academic performance of Medical undergraduate students- Cross-sectional study. Bangladesh Journal of Medical Science, 21(1), 135-139.

Tadese, M., Yeshaneh, A., & Mulu, G. B. (2022). Determinants of good academic performance among university students in Ethiopia: a cross-sectional study. BMC medical education, 22(1), 1-9.

Taguchi, Y., & Turki, T. (2022). Novel feature selection method via kernel tensor decomposition for improved multi-omics data analysis. BMC medical genomics, 15(1), 1-12.

Talsma, K., Robertson, K., Thomas, C., & Norris, K. (2021). COVID-19 Beliefs, Self-Efficacy and Academic Performance in First-year University Students: Cohort Comparison and Mediation Analysis. Frontiers in Psychology, 12, 643408-643408.

Tan, C. Y. (2024). Socioeconomic status and student learning: insights from an umbrella review. Educational Psychology Review, 36(4), 100.

Tang, H., Yang, Y., & Bao, Y. (2023). Understanding college students’ performance goals toward using open educational resources from the perspective of expectancy–value theory. Distance Education, 1-19.

Tatnall, A. (2020). Different Methodological Approaches to Considering the Adoption of ICT. Encyclopedia of Education and Information Technologies, 537-547.

Tavares, R., Vieira, R., & Pedro, L. (2017). A preliminary proposal of a conceptual educational data mining framework for science education: Scientific competences development and self-regulated learning. In 2017 international symposium on computers in education (siie) (pp. 1-6). IEEE.

Tekin, A. (2014). Early Prediction of Students' Grade Point Averages at Graduation: A Data Mining Approach. Eurasian Journal of Educational Research, 54, 207-226.

Thaher, T., Zaguia, A., Al Azwari, S., Mafarja, M. M., Chantar, H., Abuhamdah, A., Turabieh, H., Mirjalili, S., & Sheta, A. (2021). An Enhanced Evolutionary Student Performance Prediction Model Using Whale Optimization Algorithm Boosted with Sine-Cosine Mechanism. Applied Sciences, 11(21), 10237.

Thange, U., Shukla, V. K., Punhani, R., & Grobbelaar, W. (2021, January). Analyzing COVID-19 Dataset through Data Mining Tool "Orange". In 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM) (pp. 198-203). IEEE.

Tian, J., Lin, J., & Li, D. (2025). From Social to Academic: Associations and Predictions Between Different Types of Peer Relationships and Academic Performance Among College Students. Applied Sciences, 15(5), 2262.

Tian, X., Samat, N. A., & Zainal, Z. (2024). Gender Differences in Chinese EFL Learners' Comprehension When Reading Across Mediums. Editorial Board, 17(3), 32.

Tokan, M. K., & Imakulata, M. M. (2019). The effect of motivation and learning behaviour on student performance. South African Journal of Education, 39(1), 1-8.

Tong, X., & McBride, C. (2016). Reading comprehension mediates the relationship between syntactic awareness and writing composition in children: A longitudinal study. Journal of psycholinguistic research, 45, 1265-1285.

Toropova, A., Johansson, S., & Myrberg, E. (2019). The role of teacher characteristics for student performance in mathematics and student perceptions of instructional quality. Education Inquiry, 10(4), 275-299.

Torres, E. (2024). Publication: Papers on political tolerance and education. https://dash.harvard.edu/entities/publication/458cddc5-d273-41d5-b39d-5043de10b20b.

Treacy, P., O'Meara, N., & Prendergast, M. (2023). The role of expectancy-value theory in upper-secondary level students' decisions to avoid the study of advanced mathematics. Irish Educational Studies, 1-14.

Triayudi, A., & Widyarto, W. O. (2021, June). Comparison J48 and Naïve Bayes Methods in Educational Analysis. In Journal of Physics: Conference Series (Vol. 1933, No. 1, p. 012062). IOP Publishing.

Trivedi, S. (2022). Improving students' retention using machine learning: Impacts and implications. EDUCATION, 43(2), 235-258.

Trogden, B. G., Kennedy, C., & Biyani, N. K. (2023). Mapping and Making Meaning from Undergraduate Student Engagement in High-Impact Educational Practices. Innovative Higher Education, 48(1), 145-168.

Troncoso-García, A. R., Martínez-Ballesteros, M., Martínez-Álvarez, F., & Troncoso, A. (2023). A new approach based on association rules to add explainability to time series forecasting models. Information Fusion, 94, 169-180.

Tsang, G., Zhou, S. M., & Xie, X. (2020). Modelling large sparse data for feature selection: hospital admission predictions of the dementia patients using primary care electronic health records. IEEE Journal of Translational Engineering in Health and Medicine, 9, 1-13.

Tsegay, S. M., & Ashraf, M. A. (2015). The influence of senior secondary school teachers on students' performance in Gao-Kao. International Journal of Research Studies in Education, 4(4), 67-76.

Tsui, K. L., Chen, V., Jiang, W., Yang, F., & Kan, C. (2023). Data mining methods and applications. In Springer handbook of engineering statistics (pp. 797-816). London: Springer London.

Tu, X., Lv, Y., Li, X., Fang, X., Zhao, G., Lin, X., Hong, Y., Zhang, L. Y., & Stanton, B. (2009). School performance and school behaviour of children affected by acquired immune deficiency syndrome (AIDS) in China. Vulnerable Children and Youth Studies, 4(3), 199-209.

Tuffah, B. M., & Al-Jubouri, M. (2021). Relationships between Academic Stress and Socio-demographic data among Nursing Students during COVID-19 Pandemic. Kufa Journal for Nursing Sciences, 11(2), 114-123.

Udeh-Aloysius, K. N., & Achufusi, N. N. (2024). Self-efficacy and motivation as correlates of secondary school students' academic achievement in biology in Enugu South Local Government Area. International Journal of Education and Evaluation, 10(4), 237-246.

Ulfah, S. (2022). Comparison of C4.5 algorithm with the Naïve Bayes classifier algorithm for classification of nutritional status of toddler. Jurnal Disprotek, 13(1), 1-10.

Ünal, F. (2020). Data mining for student performance prediction in education. Data Mining-Methods, Applications and Systems, 28, 423-432.

Uwurukundo, M. S., Maniraho, J. F., Tusiime, M., Ndayambaje, I., & Mutarutinya, V. (2024). GeoGebra software in teaching and learning geometry of 3-dimension to improve students' performance and attitude of secondary school teachers and students. Education and Information Technologies, 29(8), 10201-10223.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Brière, N. M., Senécal, C. B., & Vallières, É. F. (1993). Academic motivation scale (AMS-C 28), college (CEGEP) version. Educational and Psychological Measurement, 52(53), 1992-1993.

Van Gestel, T., Suykens, J. A., Baesens, B., Viaene, S., Vanthienen, J., Dedene, G., Moor, B. D., & Vandewalle, J. (2004). Benchmarking least squares support vector machine classifiers. Machine learning, 54(1), 5-32.

Vargas-Ramos, J. C., Lerma, C., Guzmán-Saldaña, R., Lerma, A., Bosques-Brugada, L. E., & González-Fragoso, C. (2021). Academic Performance during the COVID-19 Pandemic and Its Relationship with Demographic Factors and Alcohol Consumption in College Students. International Journal of Environmental Research and Public Health, 19(1), 365.

Vekli, G. S., & Çalik, M. (2023). The effect of web-based biology learning environment on academic performance: a meta-analysis study. Journal of Science Education and Technology, 32(3), 365-378.

Velaj, Y., Dolezal, D., Ambros, R., Plant, C., & Motschnig, R. (2022, October). Designing a Data Science Course for Non-Computer Science Students: Practical Considerations and Findings. In 2022 IEEE Frontiers in Education Conference (FIE) (pp. 1-9). IEEE.

Villafañe, S. M., Xu, X., & Raker, J. R. (2016). Self-efficacy and academic performance in first-semester organic chemistry: testing a model of reciprocal causation. Chemistry Education Research and Practice, 17(4), 973-984.

Vollmer, A., Vollmer, M., Lang, G., Straub, A., Shavlokhova, V., Kübler, A., Gubik, S., Brands, R., Hartmann, S., & Saravi, B. (2022). Associations between periodontitis and COPD: An artificial intelligence-based analysis of NHANES III. Journal of Clinical Medicine, 11(23), 7210.

Vyasa, I. M. S., Wibawa, I. G. A., Putra, I. G. N. A. C., Agung, I. G., Kadyanan, A., ER, N. A. S., & Suputra, I. P. G. H. (2023). Apllication of Naïve Bayes Algorithm in Education Games Learn to Write Aksara Bali. Jurnal Elektronik Ilmu Komputer Udayana p-ISSN, 2301, 5373.

Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human behaviour, 104, 106189.

Wahelo, T. T., Mengistu, D. A., & Merawi, T. M. (2025). Curbing environmental problems related to deforestation and climate change: The level of secondary school students' knowledge, attitudes and determinants in Metekel Zone, Northwest Ethiopia. Geo: Geography and Environment, 12(1), e70004.

Wang, A. Y (2018). Practice and experience of the Communist Party of China in promoting educational equity during the 40 years of reform and opening up (In Chinese). Party Literature.

Wang, A., F., & Lei, X. High School. (2018). High school career planning education and its realization path under the background of new college entrance examination reform. Educational Theory and Practice (01), 33-37.

Wang, B., Yu, H., Sun, Y., Zhang, Z., & Qin, X. (2023). An effect assessment system for curriculum ideology and politics based on students' achievements in Chinese engineering education. International Journal of Advanced Computer Science and Applications, 14(1).

Wang, D., Lian, D., Xing, Y., Dong, S., Sun, X., & Yu, J. (2022). Analysis and prediction of influencing factors of college student achievement based on machine learning. Frontiers in Psychology, 13, 881859.

Wang, H., Chen, Y., Yang, X., Yu, X., Zheng, K., Lin, Q., Cheng, X. Z., & He, T. (2023). Different associations of parental involvement with children's learning of Chinese, English, and math: A three-wave longitudinal study. European Journal of Psychology of Education, 38(1), 269-285.

Wang, J. (2022). Application of C4. 5 decision tree algorithm for evaluating the college music education. Mobile information systems, 2022(1), 7442352.

Wang, L. (2022). Exploring the relationship among teacher emotional intelligence, work engagement, teacher self-efficacy, and student academic performance: A moderated mediation model. Frontiers in psychology, 12, 810559.

Wang, L., & Wang, Y. (2023). Student academic prediction model based on improved spectral clustering algorithm. Journal of Jiangsu Institute of Construction Vocational Technology (01), 45-48. (in Chinese) doi:10.19712/j.cnki.jsjyxb.2023.01.023.

Wang, L., Wang, B., Huang, H., & Zhang, Q. (2024). Research on the Teaching Reform of Data Mining and Data Analysis Based on the Concept of "Outcomes-Based Education." Educational Advances, 1(5):1-6.

Wang, X. (2021). China national college entrance examination brings new changes to high schools. Journal of Contemporary Educational Research, 5(1).

Wang, X., F. (2022). The implementation effect, multiple conflicts and coping strategies of the comprehensive reform of college entrance examination China Higher Education Research, 07, 41-47.

Watkins, D. A., & Biggs, J. B. (1996). The Chinese learner: Cultural, psychological, and contextual influences. Comparative Education Research Centre, Faculty of Education, University of Hong Kong, Pokfulam Road, Hong Kong; The Australian Council for Educational Research, Ltd., 19 Prospect Hill Road, Camberwell, Melbourne, Victoria 3124, Australia.

Watson, V. D., McClinton, T., & Watson-Lee, T. (2025). Leveraging Social Media, Images, and Artificial Intelligence to Enhance Data Access for African American Students. Black History Bulletin, 88(1), 23-25.

Webb, C. D. (2025). A Qualitative Study Examining Effective Parental Involvement Practices in a Comprehensive Community Public School District Grammar School: Perspectives of Early Childhood Teachers (Doctoral dissertation, Saint Peter's University). (Order No. 32042189).

Weedon, B. D., Liu, F., Mahmoud, W., Burden, S. J., Whaymand, L., Esser, P., Collett, J., Izadi, H., Joshi, S., Meaney, A., Delextrat, A., Kemp, S., Jones, A., & Dawes, H. (2022).

Declining fitness and physical education lessons in UK adolescents. BMJ Open Sport & Exercise Medicine, 8(1), e001165.

Wei, M. (2020). Educational equity: A comparative study of college entrance exam between China and the US from the perspective of media. Cogent Education, 7(1), 1826620.

Wen, H. T., Wu, H. Y., Liao, K. C., & Chen, W. C. (2023). JT9D Engine Thrust Estimation and Model Sensitivity Analysis Using Gradient Boosting Regression Method. Aerospace, 10(7), 639.

Whiteley, S., Porter, J., & Authority, T. E. P. (1998). Student perceptions of subject selection: Longitudinal perspectives from Queensland schools. In Australian Association for Research in Education 1998 Conference Proceedings.

Whittle, R. J., Telford, A., & Benson, A. C. (2018). Teacher's Perceptions of how they Influence Student Academic Performance in VCE Physical Education. Australian Journal of Teacher Education, 43(2), 1-25.

Wongvorachan, T., He, S., & Bulut, O. (2023). A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. Information, 14(1), 54.

Wu, D. J., Gibson, T. M., Ziegenbein, L. M., Phillis, R. W., Zehnder, C. B., Connor, E. A., & Dasgupta, N. (2024). An identity-based learning community intervention enhances the lived experience and success of first-generation college students in the biological sciences. Scientific Reports, 14(1), 10163.

Wu, Q., Fang, R., Han, B., & Pu, D. (2018). Research on student performance analysis and prediction model based on decision tree-LMBP neural network. Journal of Chengdu

Information Engineering University (03),274-280. (in Chinese) doi:10.16836/j.cnki.jcuit.2018.03.009.

Wu, S. F., Kao, C. H., Lu, Y. L., & Lien, C. J. (2022). A Method Detecting Student's Flow Construct during School Tests through Electroencephalograms (EEGs): Factors of Cognitive Load, Self-Efficacy, Difficulty, and Performance. Applied Sciences, 12(23), 12248.

Wu, T., Wei, Y., Wu, J., Yi, B., & Li, H. (2023). Logistic regression technique is comparable to complex machine learning algorithms in predicting cognitive impairment related to post intensive care syndrome. Scientific Reports, 13(1), 2485.

Xia, X. (2023). Learning behaviour mining and decision recommendation based on association rules in interactive learning environment. Interactive Learning Environments, 31(2), 593-608.

Xia, X., & Yan, J. (2021). Construction of music teaching evaluation model based on weighted naïve bayes. Scientific Programming, 2021, 1-9.

Xiao, Y., & Hu, J. (2019). Assessment of optimal pedagogical factors for Canadian ESL learner's reading literacy through artificial intelligence algorithms. International Journal of English Linguistics, 9(4), 1-14.

Xie, G., & Zhang, Y. (2020). School of golden touch? A study of school effectiveness in improving student academic performance. The Journal of Chinese Sociology, 7, 1-22.

Xiong, W., & Lai, X. (2023). A structural model of teachers' teaching competencies based on multimodal affective features of support vector machines. Applied Mathematics and Nonlinear Sciences. https://doi.org/10.2478/amns.2023.2.00201

Xu, J., & Lv, C. (2022). The influence of migrant children's identification with the college matriculation policy on their educational expectations. Frontiers in Psychology, 13, 963216.

Xu, K. (2020, March). Course selection guide, change analysis and development trend analysis of new college entrance examination reform. In International Conference on Modern Educational Technology and Innovation and Entrepreneurship (ICMETIE 2020) (pp. 279-284). Atlantis Press.

Xu, K. X. (2020, March). Course selection guide, change analysis and development trend analysis of new college entrance examination reform. In International Conference on Modern Educational Technology and Innovation and Entrepreneurship (ICMETIE 2020) (pp. 279-284). Atlantis Press.

Xu, S., & Jin, Y. (2024). Chinese adolescents' socioeconomic status and English achievement: the mediating role of parental emotional support. Humanities and Social Sciences Communications, 11(1), 1-11.

Xu, Y., Liu, H., Chen, B., Huang, S., & Zhong, C. (2023). Examining the diversity of scientific methods in college entrance chemistry examinations in China. Chemistry Education Research and Practice, 24(2), 494-508.

Xu, Z. (2023). Examining the factors influencing mathematics academic achievement in mainland China: A multilevel analysis. Oxford Review of Education, 49(3), 390-407.

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1), 1-19.

Yahdin, S., Desiani, A., Gofar, N., & Agustin, K. (2021). Application of the relief-f algorithm for feature selection in the prediction of the relevance education background with the graduate employment of the universitas sriwijaya. Computer Engineering and Applications Journal, 10(2), 71-80.

Yale, K., Nisbet, R., & Miner, G. D. (2017). Handbook of Statistical Analysis and Data Mining Applications. Elsevier.

Yan, L., & Liu, Y. (2020). An ensemble prediction model for potential student recommendation using machine learning. Symmetry, 12(5), 728.

Yang, D., Du, J., Cai, H., Wang, J., Gai, X., & Li, H. (2023, May). Application Of C4. 5 Decision Tree Algorithm For Evaluating Research-Oriented Dual-Qualified Teacher. In 2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA) (Vol. 3, pp. 131-137). IEEE.

Yang, E., & Sanborn, B. (2021). Ecological Systems Theory. In A Handbook of Theories on Designing Alignment between People and the Office Environment (pp. 101-114). Routledge.

Yang, M., Cooc, N., & Sheng, L. (2017). An investigation of cross-linguistic transfer between Chinese and English: A meta-analysis. Asian-Pacific Journal of Second and Foreign Language Education, 2(1), 1-21.

Yang, M., Cooc, N., & Sheng, L. (2017). An investigation of cross-linguistic transfer between Chinese and English: A meta-analysis. Asian-Pacific Journal of Second and Foreign Language Education, 2(1), 15.

Yang, W., Sun, R., Wang, C., Chen, J., Zhang, C., Yu, J., & Liu, H. (2023). Epidemiology of depressive disorders among youth during Gaokao to college in China: results from Hunan Normal University mental health survey. BMC psychiatry, 23(1), 1-18.

Yang, X. (2023). Research on User Demand Determination Method and Marketing Strategy for Home Broadband Business Based on Data Mining. International Journal of Data Science and Analytics.

Yang, X. Y., & Long, Y. H. (2020, May). Research on the College Entrance Examination Enrolment Reform Based on Equity Theory Under the New Situation. In 6th International Conference on Humanities and Social Science Research (ICHSSR 2020) (pp. 460-464). Atlantis Press.

Yang, Y. (2021). The evaluation of online education course performance using decision tree mining algorithm. Complexity, 2021(1), 5519647.

Yang, Y. (2025). Reciprocal relationships between parental involvement and academic performance in early adolescence: A two-year longitudinal study in China. Journal of Youth and Adolescence, 54(4), 876-889.

Yang, Y. F. (2014). Some thoughts on deepening the comprehensive reform in the field of education (Doctoral dissertation) (In Chinese).

Yao, H. M., Li, J. S., & Wang, N. (2021). BP neural network-based student performance prediction in higher education. Journal of Jilin University (Information Science Edition) (04),451-455. (in Chinese) doi:10.19292/j.cnki.jdxxp.2021.04.006.

Yen, P. Y., Sousa, K. H., & Bakken, S. (2014). Examining construct and predictive validity of the Health-IT Usability Evaluation Scale: confirmatory factor analysis and structural equation modelling results. Journal of the American Medical Informatics Association, 21(e2), e241-e248.

Yeung, J. W. K., & Xia, L. L. L. (2023). Family and Individual Contexts of Middle-School Years and Educational Performance of Youths in Middle-Aged Adulthood. International Journal of Environmental Research and Public Health, 20(4), 3279.

Yi, Q., Liu, X., Liu, C., Jia, C., & Deng, L. (2022). The Influence of Family Cultural Capital on the Subject Selection Behaviour of High School Students Under the New College Entrance Examination in Mainland China: The Mediating Role of Learning Efficacy. Journal of Higher Education Theory and Practice, 22(5), 240-254.

Yi, Y., Zhang, H., Karamti, H., Li, S., Chen, R., Yan, H., & Wang, C. (2022). The use of genetic algorithm, multikernel learning, and least-squares support vector machine for evaluating quality of teaching. Scientific Programming, 2022, 1-11.

Yokkampon, U., Chumkamon, S., Mowshowitz, A., & Hayashi, E. (2021, January). Anomaly Detection in Time Series Data Using Support Vector Machines. In Proceedings of the International Conference on Artificial Life & Robotics (ICAROB2021), Online (pp. 21-24).

Younas, M., Noor, U., Zhou, X., Menhas, R., & Qingyu, X. (2022). COVID-19, students satisfaction about e-learning and academic performance: Mediating analysis of online influencing factors. Frontiers in psychology, 13, 948061.

Young, N. A. (2020). Getting the teacher’s attention: Parent-teacher contact and teachers’ behaviour in the classroom. Social Forces, 99(2), 560-589.

Yu, H (2019). Positioning of the college entrance examination and the implementation path of content reform in the new era (In Chinese). China Examination.

Yuan, M., & Xu, D. (2024). A Practical Exploration of Teacher Training for Child Reading Education in China: A Case Study of Qinjinmuyu's" Reading Teacher Professional

Competence Certification" Program. Science Insights Education Frontiers, 25(2), 4163-4181.

Yushau, B., & Omar, M. H. (2015). Mathematics performance and its relation to English language proficiency level of bilingual Arab university students. Indian Journal of Science and Technology, 8(13), 1-15.

Zafari, M., Sadeghi-Niaraki, A., Choi, S. M., & Esmaeily, A. (2021). A practical model for the evaluation of high school student performance based on machine learning. Applied Sciences, 11(23), 11534.

Zamora, R., & Hernandez, R. (2016). The impact of organizational health on student performance in a high needs district. Journal of Studies in Education, 6(3), 149-167.

Zang, X., Wong, Y. K., & Lau, K. L. (2024). Predictors of Chinese reading in Chinese monolingual and Chinese–English bilingual children in Mainland China. Reading and Writing, 1-23.

Zang, X., Wong, Y. K., & Lau, K. L. (2024). Predictors of Chinese reading in Chinese monolingual and Chinese–English bilingual children in Mainland China. Reading and Writing, 1-23.

Zeng, X. (2025). Analyzing the Learning Behaviour Patterns of Business English Learners Using Deep Learning Technology. Systems and Soft Computing, 200259.

Zhan, S., Bray, M., Wang, D., Lykins, C., & Kwo, O. (2013). The effectiveness of private tutoring: Students' perceptions in comparison with mainstream schooling in Hong Kong. Asia Pacific Education Review, 14, 495-509.

Zhang, B., Yin, X., & Ren, Z. (2024). Can perceived social support influence academic achievement of master's students?——Evidence from a University in China. Education and Information Technologies, 29(16), 21449-21475.

Zhang, J., Liu, H., & Ni, L. (2020). A secure energy-saving communication and encrypted storage model based on RC4 for EHR. Ieee Access, 8, 38995-39012.

Zhang, J., Zhao, N., & Kong, Q. P. (2019). The relationship between math anxiety and math performance: A meta-analytic investigation. Frontiers in psychology, 10, 1613.

Zhang, K., Sun, X., Flores-Gaona, Z., Yu, C. L., Eggleston, R. L., Nickerson, N., Caruso, V. C., Tardif, T., & Kovelman, I. (2025). Cross-linguistic transfer in bilingual children's phonological and morphological awareness skills: a longitudinal perspective. Bilingualism: Language and Cognition, 28(2), 327-342.

ZHANG, L. (2024). The influence of teachers' self-efficacy on students' academic performance among secondary schools in Chongqing City, China. Region- Educational Research and Reviews, 6(7), 148-151.

Zhang, L., Mendoza, N. B., & Jiang, Y. (2025). From classroom goal structures to academic outcomes: The mediating role of expectancy-value beliefs in adolescent learning. European Journal of Psychology of Education, 40(2), 59.

Zhang, L., Yang, Y., Luo, Y., Liu, Z. Z., Jia, C. X., & Liu, X. (2022). A longitudinal study of insomnia, daytime sleepiness, and academic performance in Chinese adolescents. Behavioural sleep medicine, 20(6), 798-808.

Zhang, S. Z., Inoue, T., Zhang, D., Jin, D., & Georgiou, G. K. (2024). Intergenerational effects on children's reading comprehension in Chinese: Evidence from a 3-year longitudinal study. Scientific Studies of Reading, 28(6), 636-655.

Zhang, S.C. (2020). Difficulties and breakthrough paths in the implementation of the new college entrance examination reform. Journal of Southwest University (Social Science Edition), 46(03), 91–100+202.

Zhang, W. (2023). The History of the National College Entrance Examination Reform and its Impact on Social Mobility. LNEP, 7, 260-266.

Zhang, W., Wang, Y., & Wang, S. (2022). Predicting academic performance using tree-based machine learning models: A case study of bachelor students in an engineering department in China. Education and Information Technologies, 27(9), 13051-13066.

Zhang, X., & Qian, W. (2024). The effect of social support on academic performance among adolescents: The chain mediating roles of self-efficacy and learning engagement. PloS one, 19(12), e0311597.

Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: method review and comparison analysis. Frontiers in psychology, 12, 698490.

Zhang, Y.Q., Lu, Z.T., Jia, T.J. (2019). Did senior high school students reduce the burden under the new college entrance examination -- A survey of the examination burden of the first senior high school graduates in Zhejiang new college entrance examination. Research on Education Development, 39(12), 43–52.

Zhang, Z., Ding, Y., Zhao, X., Siye, L., Yang, N., Cai, Y., & Zhan, Y. (2023). Multireceptive field: An adaptive path aggregation graph neural framework for hyperspectral image classification. Expert Systems with Applications, 217, 119508.

Zhang, Z., WAN, S. W. Y., Chan, L. H., Kwan, P. Y. L., Tam, L. L. S., & Wan, E. W. P. (2020). Collective Teacher-Researcher Inquiry: Localizing School-Based Curriculum

Development in Diversified Hong Kong Schooling Contexts. Journal of Practitioner Research, 5(2), 3.

Zheng, X., & Li, C. (2024). Predicting students' academic performance through machine learning classifiers: A study employing the Naive Bayes Classifier (NBC). International Journal of Advanced Computer Science & Applications, 15(1).

Zheng, Y., Li, W., & Fang, L. (2019, December). Courses Arrangement Based on Optimization Model under New College Entrance Examination Reform. In 2019 2nd International Conference on Mathematics, Modelling and Simulation Technologies and Applications (MMSTA 2019) (pp. 12-15). Atlantis Press.

Zheng, Z., & Na, K. (2021). A data-driven emotion model for english learners based on machine learning. International Journal of Emerging Technologies in Learning (iJET), 16(8), 34-46.

Zhou, X., Li, J., Shen, H. T., Kitsuregawa, M., & Zhang, Y. (2005). Frontiers of WWW Research and Development--APWeb 2006: 8th Asia-Pacific Web Conference, Harbin, China, January 16-18, 2006, Proceedings (Vol. 3841). Springer.

Zhou. S. T. (2018). A cultural review of comprehensive quality evaluation incorporated into college entrance examination admission encounter impediments. China Exam [J] (04), 65-70.

Zhu, H., Li, L., & Li, H. (2022). Teacher trust in others and teacher commitment co-mediate the path from school leadership to students' reading literacy in rural Chinese schools. Frontiers in Psychology, 13, 992003.

Zhuoma, Q. (2023). Exploring an Effective Vocabulary Learning Technique for Tibetan English Language Learners in Tibet Autonomous Region: A Case Study Conducted in Qinghai Province, China. English Language Teaching, 16(3), 1-40.

Zorza, J. P., & Marino, J. (2016). Executive Functions as Predictors of School Performance and Social Relationships: Primary and Secondary School Students. The Spanish Journal of Psychology, 19, E23-E23.

Zou, Q., Liu, D., Duan, H., & Xing, L. (2016). An Evaluation Model of Academic Record of Electronic and Information Engineering Base on Complex Network Theory. In 2016 6th International Conference on Management, Education, Information and Control (MEICI 2016) (pp. 554-558). Atlantis Press.

APPENDICES

Appendix A: Student Questionnaire

Dear Students,
We invite you to take part in a crucial research study on "Performance Prediction of Compulsory Subjects and Recommendation of Subject Options for China's New College Entrance Examination". All the questionnaire information will be pushed to your personal account through the platform of Wenjuanxing (https://www.wjx.cn/), and you can open it to fill in. Your input is vital for shaping our understanding and driving positive change. Rest assured, your participation is voluntary, and all responses will be kept confidential. Your time is invaluable, and we appreciate your contribution to this important research. Thank you for your support!

Section 1: Background Information

Name:_____________ Class:_______________

1. What is your father's education?

a. Junior high school and below
b. Senior high school
c. Undergraduate degree
d. Master's degree
e. Doctoral degree

2. What is your mother's education?

a. Junior high school and below
b. Senior high school
c. Undergraduate degree
d. Master's degree
e. Doctoral degree

3. What is your annual parental income? (According to the Classification of the National Bureau of Statistics of China)

a. <10000
b. 10000-30000
c. 30000-60000
d. 60000-100000
e. >100000

4. Have you participated in private tutoring? If so, what is (are) the subject (s) of the tutoring?

a. Yes, _________
b. No

5. Have you received any form of social support? If yes, please specify the type of support received.

a. Yes, ________
b. No

Section 2: Peer relationship (Sawatzky et al, 2009)

Please rate how those statements apply to you on a 4-point scale, with only one option per question to choose. A higher score indicates a higher level of peer relationship. Thank you for your support!

	Items	1.Strongly disagree	2.A little bit of agreement	3.Quite agree	4.Totally agree
a	My friends are nice to me	⊙	⊙	⊙	⊙
b	My friends are great	⊙	⊙	⊙	⊙
c	My friends with help me if I need it	⊙	⊙	⊙	⊙
d	My friends treat me well	⊙	⊙	⊙	⊙
e	I wish I had different friends	⊙	⊙	⊙	⊙
f	I have a lot of fun with my friends	⊙	⊙	⊙	⊙
g	I have enough friends	⊙	⊙	⊙	⊙

Section 3: Parental Involvement in Students' Learning (Akello, 2020)

Thank you for your support!

1. Parents check my homework and performance.

	Items	Not at all	Rarely	Weekly	Always
a	Parents ensure that I complete my homework	⊙	⊙	⊙	⊙
b	Parents discuss with me my homework assignment	⊙	⊙	⊙	⊙
c	Parents check and sign completed homework	⊙	⊙	⊙	⊙

2. Parents attend school meetings and check my performance.

	Items	Not at all	Rarely	Sometimes	Always
a	Parents attend annual general meetings at school	⊙	⊙	⊙	⊙
b	Parents participate in parent-teacher association activities	⊙	⊙	⊙	⊙
c	Parents discuss my performance on the open day	⊙	⊙	⊙	⊙

3. Parents initiate follow-ups in school and check my performance.

	Items	Not at all	Rarely	Ferquently	Always
a	Parents create time to meet subject teachers	⊙	⊙	⊙	⊙
b	Parents request for feedback after every assessment	⊙	⊙	⊙	⊙
c	Parents make regular communication with the class teacher	⊙	⊙	⊙	⊙

4. Parents purchase supplementary materials.

	Items	Not at all	Rarely	Sometimes	Always
a	Parents purchase supplementary books	⊙	⊙	⊙	⊙
b	Parents purchase stationaries	⊙	⊙	⊙	⊙
c	Parents provide the appropriate uniform	⊙	⊙	⊙	⊙

Section 4: Motivation (Stover et al., 2012)

Academic Motivation Scale (high school version)

Please rate how those statements apply to you on a 4-point scale, with only one option per question to choose. A higher score indicates a higher level of academic motivation. Thank you for your support!

	Why do you go to school?	1.Strongly disagree	2.A little bit of agreement	3.Quite agree	4.Totally agree
1	Because I enjoy debating/ communicating/writing my ideas to others.	⊙	⊙	⊙	⊙
2	For the satisfaction I experience as I excel in my studies.	⊙	⊙	⊙	⊙
3	Because I enjoy learning new things.	⊙	⊙	⊙	⊙
4	Because I believe that without a university education, I will be ill-prepared to work in the area I like, as a university degree is not the same as a short course or a short tertiary education.	⊙	⊙	⊙	⊙
5	Because when I succeed in school I feel important.	⊙	⊙	⊙	⊙
6	Because it takes more than a high school diploma to find a well-paying job in the future.	⊙	⊙	⊙	⊙
7	Honestly, I don’t know; I really feel like I’m wasting my time at the faculty.	⊙	⊙	⊙	⊙

8	For the pleasure I experience when I participate in interesting discussions with some teachers.	⊙	⊙	⊙	⊙
9	For the satisfaction I experience as I surpass myself in my personal goals.	⊙	⊙	⊙	⊙
10	Because I like to discover new topics, related to my interests my interests, that I have never seen before.	⊙	⊙	⊙	⊙
11	Because it will allow me to enter the labour market in the field I like.	⊙	⊙	⊙	⊙
12	Because I like to have good grades and to be congratulated for that.	⊙	⊙	⊙	⊙
13	To get a more prestigious job in the future.	⊙	⊙	⊙	⊙
14	Some time ago I had reasons to go to school; however, now I wonder whether to continue or not, now I wonder whether or not to continue.	⊙	⊙	⊙	⊙
15	For the pleasure of reading about subjects that interest me.	⊙	⊙	⊙	⊙
16	For the satisfaction I feel when I manage to carry out difficult academic activities.	⊙	⊙	⊙	⊙
17	Because I enjoy increasing my knowledge of topics that appeal to me.	⊙	⊙	⊙	⊙
18	Because, in our society, it is important to go to school.	⊙	⊙	⊙	⊙
19	Because I don't want to be a failure.	⊙	⊙	⊙	⊙
20	To have a better salary in the future.	⊙	⊙	⊙	⊙
21	I can't understand why I go to school and, frankly, I couldn't care less.	⊙	⊙	⊙	⊙
22	For the satisfaction of doing something I like, for example, writing a short story in English, doing an experiment in Biology, or preparing a project or a monograph, etc.	⊙	⊙	⊙	⊙
23	Because secondary school allows me to experience personal achievement in my pursuit of excellence in my studies.	⊙	⊙	⊙	⊙
24	Because my studies allow me to continue learning many things that interest me.	⊙	⊙	⊙	⊙
25	Because I believe that my secondary education will improve my skills as a worker.	⊙	⊙	⊙	⊙

26	Because I don't want to disappoint my family.	⊙	⊙	⊙	⊙
27	I don't know; I can't understand what I'm doing at school.	⊙	⊙	⊙	⊙

Section 5: Self-Efficacy (Gafoor & Ashraf, 2016)

Academic Self-Efficacy Scale - (English Version)

Some statements concerning your beliefs about the learning are given below, five responses are given to each statement. 1. Exactly true 2. Nearly True 3. Neutral 4. Nearly False 5. Exactly false. Carefully read each statement and decide to what extent it is true in your case. Please select your response directly on the online questionnaire, choosing only one option per question. Thank you for your support!

	Items	1.Exactly True	2.Nearly True	3.Neutral	4.Nearly False	5.Exactly False
1	Irrespective of the subject, I am not competent in learning.	⊙	⊙	⊙	⊙	⊙
2	I cannot read and understand my textbooks well.	⊙	⊙	⊙	⊙	⊙
3	I sense that I am not quick to pick the points from what I read.	⊙	⊙	⊙	⊙	⊙
4	I feel unable to remember things consistently.	⊙	⊙	⊙	⊙	⊙
5	I cannot do my projects well.	⊙	⊙	⊙	⊙	⊙
6	I cannot manage time efficiently for learning.	⊙	⊙	⊙	⊙	⊙
7	I cannot arrange the help of my teachers in learning.	⊙	⊙	⊙	⊙	⊙
8	I fail to find out the necessary sources for my study.	⊙	⊙	⊙	⊙	⊙
9	I cannot arrange the help of my peers for my learning whenever I need it.	⊙	⊙	⊙	⊙	⊙
10	I fail to set higher goals in my study.	⊙	⊙	⊙	⊙	⊙
11	I cannot usually find quite a few solutions when I confront problems in my study.	⊙	⊙	⊙	⊙	⊙
12	I cannot express ideas well while attending examinations.	⊙	⊙	⊙	⊙	⊙
13	It is difficult for me to read and understand the textbooks in the English language.	⊙	⊙	⊙	⊙	⊙

14	During examinations, I cannot recollect what I have learnt.	⊙	⊙	⊙	⊙	⊙
15	Often I fail to comprehend the actual meaning of what I study.	⊙	⊙	⊙	⊙	⊙
16	If taught, I cannot prepare my class notes neatly.	⊙	⊙	⊙	⊙	⊙
17	I fail to find time for learning amid sundry chores.	⊙	⊙	⊙	⊙	⊙
18	I cannot arrange the resources of my study from my relatives and neighbours.	⊙	⊙	⊙	⊙	⊙
19	I am not assured that I have a few friends who would be helpful in my study.	⊙	⊙	⊙	⊙	⊙
20	I may not clarify doubts with my teachers while in class, even if I reach higher classes.	⊙	⊙	⊙	⊙	⊙
21	I cannot accomplish my aims in learning.	⊙	⊙	⊙	⊙	⊙
22	I cannot answer the essay-type questions well.	⊙	⊙	⊙	⊙	⊙
23	I experience that I am not weak in understanding the classes of my teachers.	⊙	⊙	⊙	⊙	⊙
24	I cannot develop the reading skills required to learn school subjects.	⊙	⊙	⊙	⊙	⊙
25	When I study a new concept, I cannot recall the related knowledge from the earlier classes.	⊙	⊙	⊙	⊙	⊙
26	I cannot utilize the available library facility for my study.	⊙	⊙	⊙	⊙	⊙
27	I observe that I fail to prepare my seminars and assignments in time.	⊙	⊙	⊙	⊙	⊙
28	If I miss some classes for some reason, I cannot compensate for the loss fairly well.	⊙	⊙	⊙	⊙	⊙
29	I consider that I failed to develop a healthy relationship with my teachers.	⊙	⊙	⊙	⊙	⊙
30	I am not confident that I can perform well in competitive examinations.	⊙	⊙	⊙	⊙	⊙

31	I cannot deal efficiently with the unexpected problems in my study.	⊙	⊙	⊙	⊙	⊙
32	I cannot be calm at exam time as I am conscious of my ability to learn.	⊙	⊙	⊙	⊙	⊙
33	I cannot complete the homework myself without any help from guidebooks, previous notes, etc.	⊙	⊙	⊙	⊙	⊙
34	I cannot usually handle disturbing situations in the study.	⊙	⊙	⊙	⊙	⊙
35	If a sudden test is conducted for us without prior notice, I cannot answer it well.	⊙	⊙	⊙	⊙	⊙
36	If I try, I cannot become one of the good grade holders.	⊙	⊙	⊙	⊙	⊙
37	I cannot answer the questions which teachers ask me.	⊙	⊙	⊙	⊙	⊙
38	I cannot score well in the short answer type questions.	⊙	⊙	⊙	⊙	⊙
39	I cannot accomplish challenging tasks and problems in my study.	⊙	⊙	⊙	⊙	⊙
40	However twisted the question is, I cannot answer them.	⊙	⊙	⊙	⊙	⊙

Section 6: Test Anxiety (Putwain et al., 2020)

Multidimensional Test Anxiety Scale

Before a test/examination, rate the relevance of each item to that construct on a scale from 1 (Strongly disagree) to 5 (Totally agree). Please select your response directly on the online questionnaire, choosing only one option per question. Thank you for your support!

	Items	1.Strongly disagree	2.A little bit of agreement	3.Neutral	4.Quite agree	5.Totally agree
1	Before a test/exam, I worry that I will fail.	⊙	⊙	⊙	⊙	⊙
2	I am afraid of writing the wrong answer during a test/exam.	⊙	⊙	⊙	⊙	⊙
3	After a test/exam, I worry that I have failed.	⊙	⊙	⊙	⊙	⊙
4	After taking a test/exam, I worry that I gave the wrong answers.	⊙	⊙	⊙	⊙	⊙

5	I forget previously known material before taking a test/exam.	⊙	⊙	⊙	⊙	⊙
6	I forget facts I have learnt during tests/exams.	⊙	⊙	⊙	⊙	⊙
7	During tests/exams, I forget things that I have learnt.	⊙	⊙	⊙	⊙	⊙
8	During tests/exams, I find it hard to concentrate.	⊙	⊙	⊙	⊙	⊙
9	Even when I have prepared for a test/exam I feel nervous about it.	⊙	⊙	⊙	⊙	⊙
10	I feel tense before taking a test/exam.	⊙	⊙	⊙	⊙	⊙
11	Just before I take a test/exam, I feel panicky.	⊙	⊙	⊙	⊙	⊙
12	Before a test/exam, I feel nervous.	⊙	⊙	⊙	⊙	⊙
13	Before I take a test/ exam my hand trembles.	⊙	⊙	⊙	⊙	⊙
14	My heart races when I take a test/exam.	⊙	⊙	⊙	⊙	⊙
15	During a test/exam, I experience stomach discomfort.	⊙	⊙	⊙	⊙	⊙
16	My hand shakes while I am taking a test/exam.	⊙	⊙	⊙	⊙	⊙

Appendix B: Teacher Questionnaire

Dear Teachers,

We cordially invite you to participate in our research study titled "Performance Prediction of Compulsory Subjects and Recommendation of Subject Options for China's New College Entrance Examination". Your valuable insights are crucial to our understanding of the factors influencing students' performance and the optimization of subject options. To facilitate your participation, we will be sending the questionnaire via Wenjuanxing (https://www.wjx.cn/), an online survey platform. Rest assured that all information provided will be kept strictly confidential and used solely for research purposes. Your time and contribution are greatly appreciated. Please click on the link to access the survey and share your valuable opinions. Thank you for your support!

Section 1: Background Information

Name:______________ Teaching Class:________________

1. What is your education level?

a. Undergraduate degree
b. Master's degree
c. Doctoral degree

2. What is your qualification?

a. Secondary school teacher II
b. Secondary school teacher I
c. Secondary school associate senior teacher
d. Secondary school full senior teacher

3. What is the most common teaching method you use in your daily teaching？

a. Traditional teaching method
b. Smart classroom teaching method
c. Both

Section 2: Teacher's Efficacy

Teachers' Sense of Efficacy Scale (TSES) (Dupuis et al., 2020)

Please select your response directly on the online questionnaire, choosing only one option per question. 1.Limited: Lacks knowledge, skills, or confidence to effectively address issues. 2.Basic: Has fundamental understanding and can apply basic strategies. 3.Competent: Proficient in applying a range of strategies with some consistency. 4.Skilled: Highly proficient, using advanced strategies and techniques. 5.Expert: Deep understanding, innovative solutions, and ability to mentor and contribute to the field. Thank you for your support!

	Items	1.Limited	2. Basic	3.Competent	4.Skilled	5.Expert

1	How much can you do to control disruptive behaviour in the classroom?	⊙	⊙	⊙	⊙	⊙
2	How much can you do to motivate students who show low interest in schoolwork?	⊙	⊙	⊙	⊙	⊙
3	How much can you do to get students to believe they can do well in schoolwork?	⊙	⊙	⊙	⊙	⊙
4	How much can you do to help your students value learning?	⊙	⊙	⊙	⊙	⊙
5	To what extent can you craft good questions for your students?	⊙	⊙	⊙	⊙	⊙
6	How much can you do to get children to follow classroom rules?	⊙	⊙	⊙	⊙	⊙
7	How much can you do to calm a student who is disruptive or noisy?	⊙	⊙	⊙	⊙	⊙
8	How well can you establish a classroom management system with each group of students?	⊙	⊙	⊙	⊙	⊙
9	How much can you use a variety of assessment strategies?	⊙	⊙	⊙	⊙	⊙
10	To what extent can you provide an alternative explanation for example when students are confused?	⊙	⊙	⊙	⊙	⊙
11	How much can you assist families in helping their children do well in school?	⊙	⊙	⊙	⊙	⊙
12	How well can you implement alternative strategies in your classroom?	⊙	⊙	⊙	⊙	⊙

Appendix C: Research Ethics Letter

Research Ethics Letter

Deputy Vice Chancellor's Office
(Research and Innovation)
Human Research Ethics Committee
(Non Medical)
Tel: 082 581222/1223
Fax: 082 665115

UNIVERSITI MALAYSIA SARAWAK
94300 Kota Samarahan

MEMORANDUM

Reference : UNIMAS/TNC(PI)/09-65/02 Jld.2 (28)

To : Prof Dr Chen Chwen Jen
Faculty of Cognitive Sciences and Human Development

From : Chair
Human Research Ethics Committee (Non Medical)

Date : 17 January 2025

Subject : **Research Ethics Approval for Non-Medical Research on Humans**

With reference to the above, I would like to inform you that your application for research ethics clearance was discussed in the 13th Human Research Ethics Committee (Non-medical) meeting 4/2024 on 2 December 2024. Your application for research ethics has been approved.

Title of project :	Performance Prediction of Compulsory Subjects and Recommendation of Subject Options for China's New College Entrance Examination
No. Ethics Approval :	HREC(NM)/2023 (2)/28
Principal Investigator :	Prof Dr Chen Chwen Jen (Universiti Malaysia Sarawak)
Co-researcher(s):	Wang Long (Universiti Malaysia Sarawak)

Thank you.

Yours sincerely,

Professor Dr Ting Su Hie

c.c :Deputy Vice Chancellor (Research & Innovation)
:Director, UNIMAS Publisher
:Deputy Dean, Faculty of Cognitive Sciences and Human Development

Appendix D: CONSENT FORM

CONSENT FORM

Performance Prediction of Compulsory Subjects and Recommendation of Subject Options for China's New College Entrance Examination (Wang Long)

If you agree to participate in the study, please tick the boxes.

	Yes
1. I have read the information sheet dated _____03/09/2024_____ for this study. I have had the opportunity to consider the information, ask questions, and have had these answered satisfactorily.	✓
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my legal rights being affected.	✓
3. (If appropriate) *I agree to the [interview/focus group] being audio/video recorded.*	
4. (If appropriate) *I agree to be contacted about ethically approved studies for which I may be suitable. I understand that agreeing to be contacted does not oblige me to participate in any further studies.*	
5. I understand that all data I provide will be kept confidential and stored securely at ________the researchers' password laptop________________ (state location). Only the researchers will have access to this.	✓
6. I understand that my data will remain anonymous at all times, including when the results of the study are disseminated in conferences, published articles, and/or ____________________ (state other platforms, if any).	✓
7. I have been provided with details of whom I should contact about the study.	✓
8. I agree to take part in this study.	✓

Wang Yongqing	*03/09/2024*	
Name of the headmaster	*Date*	*Signature*

Appendix E: The Pearson's Correlation Coefficient of Exact Relationship among Chinese, Mathematics Performance

		1 Chinese	1 Mathematics	1 English	2 Chinese	2 Mathematics	2 English	3 Chinese	3 Mathematics	3 English	4 Chinese	4 Mathematics	4 English	5 Chinese	5 Mathematics	5 English
1 Chinese	Pearson Correlation	1	.124**	.308**	.280**	.182**	.308**	.275**	.156**	.309**	.217**	.109**	.277**	.233**	.111**	.249**
	Significance (2-tailed)		0	0	0	0	0	0	0	0	0	0	0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
1 Mathematics	Pearson Correlation	.124**	1	.220**	.137**	.462**	.225**	.155**	.436**	.251**	.135**	.340**	.160**	.094**	.352**	.129**
	Significance (2-tailed)	0		0	0	0	0	0	0	0	0	0	0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
1 English	Pearson Correlation	.308**	.220**	1	.333**	.330**	.559**	.304**	.338**	.610**	.299**	.263**	.468**	.243**	.271**	.488**
	Significance (2-tailed)	0	0		0	0	0	0	0	0	0	0	0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
2 Chinese	Pearson Correlation	.280**	.137**	.333**	1	.264**	.377**	.324**	.257**	.388**	.312**	.246**	.312**	.330**	.178**	.334**
	Significance (2-tailed)	0	0	0		0	0	0	0	0	0	0	0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
2 Mathematics	Pearson Correlation	.182**	.462**	.330**	.264**	1	.402**	.255**	.643**	.403**	.248**	.503**	.322**	.154**	.540**	.333**
	Significance (2-tailed)	0	0	0	0		0	0	0	0	0	0	0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
2 English	Pearson Correlation	.308**	.225**	.559**	.377**	.402**	1	.323**	.362**	.709**	.346**	.298**	.595**	.297**	.292**	.591**
	Significance (2-tailed)	0	0	0	0	0		0	0	0	0	0	0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
3 Chinese	Pearson Correlation	.275**	.155**	.304**	.324**	.255**	.323**	1	.220**	.311**	.312**	.214**	.314**	.301**	.229**	.307**
	Significance (2-tailed)	0	0	0	0	0	0		0	0	0	0	0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
3 Mathematics	Pearson Correlation	.156**	.436**	.338**	.257**	.643**	.362**	.220**	1	.368**	.231**	.520**	.301**	.135**	.534**	.333**
	Significance (2-tailed)	0	0	0	0	0	0	0		0	0	0	0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
3 English	Pearson Correlation	.309**	.251**	.610**	.388**	.403**	.709**	.311**	.368**	1	.353**	.350**	.612**	.306**	.351**	.604**
	Significance (2-tailed)	0	0	0	0	0	0	0	0		0	0	0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
4 Chinese	Pearson Correlation	.217**	.135**	.299**	.312**	.248**	.346**	.312**	.231**	.353**	1	.267**	.416**	.366**	.282**	.406**
	Significance (2-tailed)	0	0	0	0	0	0	0	0	0		0	0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
4 Mathematics	Pearson Correlation	.109**	.340**	.263**	.246**	.503**	.298**	.214**	.520**	.350**	.267**	1	.365**	.211**	.620**	.367**
	Significance (2-tailed)	0	0	0	0	0	0	0	0	0	0		0	0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
4 English	Pearson Correlation	.277**	.160**	.468**	.312**	.322**	.595**	.314**	.301**	.612**	.416**	.365**	1	.377**	.406**	.708**
	Significance (2-tailed)	0	0	0	0	0	0	0	0	0	0	0		0	0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
5 Chinese	Pearson Correlation	.233**	.094**	.243**	.330**	.154**	.297**	.301**	.135**	.306**	.366**	.211**	.377**	1	.197**	.404**
	Significance (2-tailed)	0	0	0	0	0	0	0	0	0	0	0	0		0	0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
5 Mathematics	Pearson Correlation	.111**	.352**	.271**	.178**	.540**	.292**	.229**	.534**	.351**	.282**	.620**	.406**	.197**	1	.377**
	Significance (2-tailed)	0	0	0	0	0	0	0	0	0	0	0	0	0		0
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
5 English	Pearson Correlation	.249**	.129**	.488**	.334**	.333**	.591**	.307**	.333**	.604**	.406**	.367**	.708**	.404**	.377**	1
	Significance (2-tailed)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	N	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762	6762
** Correlation is significant at the 0.01 level (2-tailed).																